Evaluation of the Health Care Innovation Awards, Round 2: Final Report

September 2020

Boyd Gilman, Danielle Whicher, Randall Brown, Nancy McCall, Stacy Dale, Laurie Felland, Robert Schmitz, John Schurrer, Lauren Vollmer Forrow, Mark Flick, Jim Derzon, and Rizwaan Lakhani

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05

COR and Evaluation C-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

Mathematica
Progress Together

Evaluation of the Health Care Innovation Awards, Round 2: Final Report

Part I: Synthesis and Summary of Main Findings

This page has been left blank for double-sided copying.

PART I CONTENTS

TABLES

FIGURES

HCIA R2 AWARDEE ACRONYMS/ABBREVIATIONS

Awardee	Acronym/abbreviation
Altarum Institute	Altarum
American College of Cardiology Foundation	ACCF
Amerigroup	Amerigroup
Association of American Medical Colleges	AAMC
Avera Health	Avera
The Board of Trustees at The University of Illinois, Chicago	UIC
Boston Medical Center	BMC
CareChoice Cooperative	CCC
Catholic Health Initiatives Iowa Corporation	CHIIC
Children's Home Society of Florida	CHS
City of Mesa Fire and Medical Department	Mesa
Clifford W. Beers Guidance Clinic	Clifford Beers
Community Care of North Carolina	CCNC
Detroit Medical Center	DMC
Four Seasons Compassion for Life	FSCL
The Fund for Public Health in New York	FPHNY
Icahn School of Medicine at Mount Sinai	Icahn
Johns Hopkins University	Hopkins
Montefiore Medical Center	Montefiore
National Association of Children's Hospitals and Related Institutions	NACHRI
National Health Care for the Homeless Council	NHCHC
Nebraska Medicine	NMC
New York City Health + Hospitals	NYC H+H
Northwell Health	Northwell
Regents of the University of California at San Diego	UCSD
Regents of the University of California at San Francisco	UCSF
Regents of the University of Michigan	UMich
Seattle Children's Hospital	SCH
Trustees of Columbia University in the City of New York	Columbia
University Hospitals Cleveland Medical Center	UHCMC
University of Kansas Health System	UKS
University of New Mexico	UNM
University of North Carolina at Chapel Hill	UNC
Ventura County Health Care Agency	Ventura
Village Center for Care	VCC
Washington University School of Medicine	Wash U
Wisconsin Department of Health Services	WI DHS
Yale University	Yale

This page has been left blank for double-sided copying.

EXECUTIVE SUMMARY

Introduction

Building off the Health Care Innovation Awards, Round 1 (HCIA R1), in which 107 awardees received over $826 million in funding, Round 2 of the Health Care Innovation Awards (HCIA R2) awarded nearly $322 million to 39 new organizations as three-year cooperative agreements. Both rounds were funded to implement models designed to improve the quality of care and health and lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries. While HCIA R1 focused on high-risk/high-opportunity populations and workforce development, HCIA R2 focused on service delivery models along with the design of corresponding new payment models. Beginning in September 2014, the HCIA R2 organizations implemented their proposed models. The programs varied widely in the interventions provided, the populations served, the types of organizations involved, the number of individuals enrolled, and the models proposed to pay for intervention services (Table ES.1).

The findings in this report—the fourth and final evaluation report of the HCIA R2 cooperative agreements—focus on (1) evidence of program impacts on health care service use and costs, (2) factors associated with evidence of favorable impacts, and (3) awardees' experiences sustaining programs and implementing payment models after the end of the awards.

Overall, of the 38 awardees that remained in the program, 19 met the criteria required for producing valid estimates of program impacts. For the other 19 awardees, the evaluation produced descriptive information only. Of the 19 programs with rigorous impact evaluations, four programs demonstrated statistically significant impacts: two on service use and expenditure measures and two on service use measures only (data on Medicaid spending for these two awardees were not available). Six programs had mixed results with either favorable but non-statistically significant results or a mix of favorable and unfavorable results. Nine programs had mainly unfavorable results.

Despite the considerable infusion of resources in HCIA R2, there is little evidence showing the potential for savings to offset such an investment. In addition, although the purpose of the cooperative agreement was to support and facilitate investigator-defined initiatives in service delivery and payment models, the cooperative agreement approach to identifying promising models does not lend itself to evaluating the impact of implementation of changes in payment policy.

This report contains two parts. **Part I** synthesizes the main impact findings of the evaluations. **Part II** includes the 38 awardee-specific evaluation reports, which formed the basis for the synthesis.[1]

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the programs. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Table ES.1. Characteristics of HCIA R2 programs, by target population

Awardee	Program description
Youth with complex medical conditions	
BMC	Team-based care to enhance outcomes for children with medical complexity in medical home-like settings
NACHRI	National learning collaborative to improve care coordination and management for children with medical complexity
SCH	Care management and provider education to reduce unnecessary or redundant services for children with disabilities
UIC	Coordination of nonmedical, medical, and mental health services to address social determinants for health for children with chronic medical conditions
WI DHS	Care management and care coordination to improve care for children with medical complexity and high-resource utilization
High-risk chronic conditions	
DMC	Patient-centered medical home clinics established adjacent to EDs to increase availability of primary care for individuals who arrived at the ED and required non-urgent care
FSCL	Team-based and patient-centered palliative care to improve quality of life for beneficiaries with life-limiting illnesses
NMC	Remote patient monitoring for beneficiaries with diabetes for 90 days after a hospital discharge to improve self-management
Northwell	Care coordination and patient education to improve decision making and quality of life for beneficiaries with advanced chronic kidney disease
NYC H+H	Care management and 90-day care coordination services for beneficiaries with ambulatory sensitive conditions who visited the ED to improve the linkage to preventive health services
UKS	Evidence-based protocols, provider education, telemedicine, and care management to improve outcomes for beneficiaries experiencing a heart attack or stroke
UCSD	Patient education provided by health coaches to beneficiaries at elevated risk of cardiovascular disease to reduce the incidence of health attacks and strokes
UHCMC	Care management and palliative care to enhance quality and patient experience for beneficiaries with complex cancers
Low-risk chronic conditions	
ACCF	Evidence-based decision-support tools to improve the appropriateness and quality of care for beneficiaries with stable ischemic heart disease
CCNC	Pharmacy-based medication management to improve medication adherence and self-management for beneficiaries with chronic conditions
CHIIC	Expanded population health activities, such as health coaching at primary care clinics, for rural, low-income communities with high rates of diabetes, obesity, and disability
FPHNY	Care coordination, tele-mentoring and peer navigation for beneficiaries with hepatitis C to improve treatment adherence
Ventura	Improved treatment guideline adherence and community-based care coordination for beneficiaries with COPD
VCC	Technology-based self-management tool to increase patient activation and treatment adherence for people living with HIV

Table ES.1 (*continued*)

Awardee	Program description
Behavioral health and cognitive disorders	
Amerigroup	Health coaching services to help youth transitioning out of foster care to access, coordinate, and manage health and social services on their own
Clifford Beers	Care management services to help children with complex needs and their families manage, coordinate, and integrate behavioral and physical health services and social supports
Johns Hopkins	Care management program for adults with Alzheimer's or other dementia-related disease and their caregivers to improve health outcomes and prevent or delay institutionalization
Montefiore	Behavioral health screening and treatment services in primary care settings to improve health outcomes and reduce the cost of care
UCSF	Care management and caregiver support for participants with dementia to prevent use of emergency-related health care services and keep patients in the community longer
Acute and subacute care	
Avera	Telehealth long-term care services for short- and long-term residents of participating rural nursing facilities to reduce unnecessary transfers to EDs and hospitals
CCC	Web-based application to improve the care and safety of SNF patients who transitioned to the community and reduce readmissions and costs of care for these patients
Icahn	Mobile team-based services to address acute and post-acute care needs in the patient's home
Mesa	On-site treatment through mobile units for low-acuity patients who used the 911 system and the ED and care transition services to patients at high risk of readmission
NHCHC	Medical respite care for homeless Medicaid and Medicare beneficiaries following discharge from a hospital or other community-based setting
UMich	Pre-operative care for patients undergoing major abdominal surgery by enrolling them in a patient education and physical activity monitoring program to reduce surgical complications
UNM	Telehealth consultations for patients who presented at a participating hospital's ED with a neuro-emergent condition
Primary and preventive care	
AAMC	Decision support templates to help primary care physicians seeking guidance about patients' treatment and to increase effectiveness of specialty care referral process
Altarum	Preventive oral health and dental services in primary care settings for Medicaid beneficiaries up to age 17
CHS	Community- and school-based model to provide comprehensive services that promote health and well-being for residents and students living in Pine Hills, Florida
Columbia	Chronic disease management model and tablet-based software suite within pediatric dentistry to reduce early childhood caries in low-income children
U NC	Care delivery model based on patient education, shared decision-making tools, and nurse patient navigators to treat new onset of lower back pain
Wash U	Contraceptive counseling and family planning services for women who were at the highest risk for unintended pregnancy
Yale	Community-based, short-term care management services to improve health of elders with impaired mobility to prevent injuries and associated use of emergency services from a fall

Note: The acronyms used for awardees are defined on page ix.

Table ES.1 (*continued*)

CHIP = Children's Health insurance Program; CKD = chronic kidney disease; CMC = children with complex medical conditions; COPD = chronic obstructive pulmonary disease; ED = emergency department; EMR = electronic medical record; FFS = fee-for-service; SNF = skilled nursing facility.

Challenges and methods of measuring program impacts

Although allowing organizations to build on their own experiences to implement novel health care delivery and payment models enabled the Centers for Medicare & Medicaid Services (CMS) to test a wide range of models simultaneously, it introduced a unique set of evaluation challenges. To overcome these challenges and produce valid estimates of program impact, each awardee evaluation had to meet three criteria:

1. A credible comparison group had to be identifiable using claims data, or from a randomized design.
2. The study population had to be large enough to detect a 20 percent effect of the program on total expenditures, number of hospitalizations, or number of ED visits.
3. The key outcomes that the program expected to affect had to be measurable using Medicare or Medicaid claims.

Of the 38 evaluated programs, 19 met these criteria, making it possible to estimate their impacts on at least one of the four core outcome measures—Medicare and Medicaid expenditures, hospitalizations, ED visits, and hospital readmissions within 30 days (Table ES.2). Of the 19 impact evaluations, 9 were based on Medicare beneficiaries only, 6 were based on Medicaid beneficiaries only, and 4 included both Medicare and Medicaid beneficiaries. Although many programs served both Medicare and Medicaid beneficiaries, the impact analyses focused on the payer group that represented the primary target population for the intervention. Participants who were dually eligible for both Medicare and Medicaid were included in the Medicare population. Four programs enrolled sizable portions of both Medicare and Medicaid beneficiaries and therefore, the evaluations of these programs estimated impacts separately for the two payer populations.

For the other 19 awardees, there were significant concerns about conducting a rigorous impact analysis based on one or more of the three criteria. Of the 19 programs without impact estimates, 10 served primarily Medicare beneficiaries, 8 served primarily Medicaid enrollees, and 1 served Medicare and Medicaid beneficiaries. Programs that (1) targeted chronic or complex conditions, (2) served all eligible individuals who received care at a participating site, (3) involved an organization-level (rather than patient-level) intervention, and (4) included a health information technology and decision support component were more likely to be evaluable than those that did not.

Table ES.2. Awardees with and without estimates of program impacts, by payer

Payer	Awardees with impact estimates	Awardees without impact estimates
Medicare only	AAMC, Avera, CCC, CCNC, CHIIC, FSCL, NMC, UCSF, UKS	ACCF, Hopkins, Icahn, Northwell, UHCMC, UMich, UNC, UNM, Ventura, Yale
Medicaid only	Altarum, Amerigroup, Montefiore, NACHRI, SCH, UIC	BMC, CHS, Clifford Beers, Columbia, DMC, NHCHC, Wash U, WI DHS
Medicare and Medicaid	FPHNY, Mesa, NYC H+H, VCC	UCSD

To determine which of the 19 evaluable programs had evidence of favorable effects on at least one of the core outcomes, three assessment rules (focusing on the study sample, evaluation follow-up period, and impact results) were applied to all 19 evaluations.

1. **Study sample**. If the results were favorable for the full study sample, those results were used to assess program impact. If the results were not favorable for the full sample but were favorable for a subgroup of relevance to the intervention, the program intervention was identified as favorable for that subgroup. The study presented impact estimates for the same subgroup across all outcomes.
2. **Evaluation follow-up period**. If the results were similar across each 12-month follow-up period, the cumulative results over the full follow-up period were used to assess program impact. If the results differed across follow-up periods, the follow-up period that was most consistent with the awardee's theory of action was used instead.
3. **Impact results**. Programs that had at least one favorable and statistically significant impact estimate on a core outcome that was consistent with the awardee's theory of action were identified as having evidence of a favorable impact. However, this favorable assessment was rejected if impact estimates for any other core outcomes were adverse and either large or statistically significant.

It is important to note that, although the rules for assessing impacts were the same, each awardee was assessed separately and the rules led to focusing on different outcomes, follow-up periods, and subgroups for different awardees. The goal was to identify those programs that had convincing evidence of favorable impacts on one or more of the core outcomes even if those programs did not have statistically significant findings for all outcomes or for all enrollees over the full program period.

Programs with evidence of favorable impacts on core outcomes

Under these assessment rules, 4 of the 19 programs with impact evaluations had evidence of favorable effects on one or more of the core outcomes for at least one 12-month follow-up period (Table ES.3). For three of the 4 programs (Avera, NYC H+H, and UIC), the favorable effects

were limited to a subgroup of beneficiaries expected to receive the greatest benefit from the program intervention.

Three of the 4 programs (Avera, Montefiore, and NYC H+H) had statistically significant estimated reductions in ED visits, ranging from 8 to 14 percent. One awardee (NYC H+H) also had a statistically significant estimated reduction in hospital admissions of 6 percent; the other three had favorable but not statistically significant estimated impacts on this outcome. These estimated effects on major cost drivers led to statistically significant estimated reductions in spending among important subgroups of eligible beneficiaries in the two programs with available expenditure data (UIC and Avera). Given the statistically significant estimated reductions in ED visits and hospital admissions for NYC H+H, and the large estimated effects on ED visits and hospitalizations for Montefiore, it is likely that these two programs also reduced total cost of care for their Medicaid participants.

1. **Avera's eLongTermCare (eLTC) program** primarily served Medicare beneficiaries. It provided geriatric telehealth consults, timely access to specialists, and training to help Avera's nursing staff address complex and/or urgent health care needs of both short- and long-term residents in nursing facilities, especially in isolated rural areas. Statistically significant estimates showed reduced Medicare spending and ED service use, but only among the long-term care beneficiary subgroup. The impacts were estimated over all nursing facilities residents, regardless of whether they received eLTC services, because the program was a facility-level intervention.
2. **Montefiore's Behavioral Health Integration Program (BHIP)** mainly served Medicaid enrollees. It provided integrated behavioral health services in the primary care setting for adults and children with behavioral health disorders. The BHIP led to a statistically significant estimated reduction in ED service use and a large but not significant estimated reduction in hospital admissions (expenditure data for Medicaid-eligible children were not available for this awardee). Program impacts were estimated over participants only.
3. **NYC H+H's Emergency Department Care Management (ED care management) program** served both Medicare beneficiaries and Medicaid enrollees. It provided care management in the ED and transitional care coordination for up to 90 days after discharge to help patients keep their follow-up physician appointments. The statistically significant estimates showed the program reduced ED service use and hospital admissions among Medicaid beneficiaries who enrolled during the first nine months of the program, and were substantially sicker on average than those enrolling later (expenditure data for Medicaid enrollees were not available for this awardee). Estimated effects for Medicare enrollees were smaller and not statistically significant. The evaluation could not replicate NYC H+H's program enrollment criteria using claims data because the awardee selectively enrolled patients based on clinicians' judgment, so the evaluation included all eligible Medicaid beneficiaries to avoid selection bias.

4. **UIC's Coordination of Health Care for Complex Kids (CHECK) program** mainly served Medicaid enrollees. It coordinated medical, nonmedical, and mental health services for children with complex medical conditions. The large and statistically significant impact estimate indicates that among higher-risk children UIC reduced Medicaid spending substantially, as a result of estimated reductions (that were sizeable but not statistically significant) in hospital inpatient and ED service use. Estimated effects on the likelihood of hospital admission and the likelihood of an ED visit (not shown here) were statistically significant. The evaluation was based on a randomized controlled trial design, with impacts estimated over all randomly assigned beneficiaries.

Table ES.3. Programs with favorable impacts on core outcomes

Awardee	Treatment group sample size	ED visits (per 1,000 beneficiaries)		Hospital admissions (per 1,000 beneficiaries)		Total expenditures ($ PBPM)	
		Impact estimate	Percentage impact	Impact estimate	Percentage impact	Impact estimate	Percentage impact
Avera[a]	7,194	-73***	-9%	-10	-1%	-$73*	-4%
Montefiore[b]	1,758	-177*	-14%	-56	-18%	NA	NA
NYC H+H[b,c]	9,747	-213***	-8%	-57*	-6%	NA	NA
UIC[d]	821	-56	-6%	-17	-15%	-$44**	-21%

[a] Favorable impacts were limited to long-term care residents.

[b] Total expenditures data were not available for Medicaid enrollees.

[c] Favorable impacts were limited to Medicaid beneficiaries who enrolled during the first nine months of program. Results for Medicare beneficiaries were small and not statistically significant.

[d] Favorable impacts were limited to higher-risk children. The treatment group sample size reflects the number of higher-risk children who were randomized to the intervention arm of the randomized control trial implemented to evaluate the program. The size of the estimated impact on total expenditures was sensitive to high-cost outliers. When adjusted for the high-cost outliers, the impacts on spending are smaller (7 percent) and not statistically significant.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month; NA = not available.

Although the impact assessment found only these four programs to have strong evidence of favorable effects on health care expenditures and service use, six others had one or more favorable estimates but did not meet one of the three assessment rules described earlier to be classified as effective. Five programs had estimated reductions in health care expenditures and service use, but none of the estimates were statistically significant (UCSF, CHIIC, NMC, and Amerigroup). Although they might have had favorable effects, the evidence was too weak (and in some cases inconsistent across outcomes) to classify them as effective. Two programs (CCC and NACHRI) had statistically significant favorable estimated effects on ED visits, but large and unfavorable estimates for other core outcomes, which made it untenable to classify these awardees as effective.

Three other programs (FSCL, Mesa, and NACHRI) led to statistically significant *increases* in either service use or total expenditures. The likely reasons for these increases differed. For FSCL, the increases in total expenditures were likely due to program enrollees receiving more intensive hospice services for a longer period. For Mesa, the increases in ED visits were likely the result of challenges associated with changing the behavior of individuals who are frequent ED users; short-term decreases in ED use were not sustained over time. Finally, for NACHRI, increases in hospitalizations may be due to program participants having more complex needs compared to individuals in the comparison group.

The remaining programs funded under HCIA R2 had estimates that were small and suggested it was highly unlikely that they may have been effective (9), or were not evaluable in a way that could produce credible impact estimates (19).

Factors associated with evidence of favorable impacts

The evaluation looked across 14 of the 19 programs with impact estimates and identified seven program features—including 3 measures of the type of intervention and 4 measures of awardee or program characteristics—that were associated with more favorable estimated program impacts on at least one of the four core outcomes.[2] These 7 features were more prevalent among the 4 programs classified as having evidence of favorable impacts than they were among the other programs. The median impacts for programs with these features were also more favorable than the median impacts for programs without the feature, for one or more of the core outcomes.

The intervention components associated with more favorable estimated impacts were integration of behavioral health with physical health care, using telemedicine, and relying on health IT to improve communications across providers. The program design or awardee characteristics that were associated with more favorable estimated impacts on one or more outcomes included having prior experience implementing similar programs, serving a socially fragmented population, focusing on individual patient care rather than transforming provider practice, and using nonclinical staff as frontline providers of the intervention. However, the difference between awardees that did or did not use nonclinical frontline staff was less pronounced than those found for other characteristics.

All four programs with evidence of favorable effects served socially fragile populations—meaning people beset with social problems such as poverty, isolation, lack of transportation, or unstable housing—in addition to their physical problems, and had experience implementing their intervention before receiving their HCIA R2 award. A few of the programs without evidence of favorable impacts also had these two characteristics. However, only the four effective programs also met the nonmedical needs of their socially fragile participants by having either a behavioral health component or by using nonclinical staff to work directly with the participants. Behavioral

[2] This analysis excluded five awardees without evidence of favorable effects: four (AAMC, Altarum, Amerigroup, and FSCL) because their interventions were not expected to affect the outcomes in the two- to three-year follow-up period of the study and one (Mesa) because the impact estimates appeared to suffer from selection bias.

health interventions can help these patients address the nonphysical problems they often face, such as depression or anxiety or substance abuse, while nonclinical staff such as community health workers and social workers are experienced in helping patients identify and access available social services and supports. Of the four programs with evidence of favorable effects, three also relied on telehealth and health information technology in delivering their interventions and three focused on meeting patient's needs rather than on changing providers' behavior. Although these features are neither necessary nor sufficient for an innovative program to have favorable effects, they suggest that programs with such characteristics might be more likely to succeed in improving care and lowering costs within a relatively short period.

Findings from program sustainability and payment models

Awardees made considerable effort to continue their programs after the award period, and many were able to sustain at least parts of their programs. Eleven awardees reported they were sustaining their entire programs (usually with minor modifications), while 18 reported sustaining only certain program components or limiting the number of people the program would reach. Three factors helped awardees obtain funds for their programs: (1) gaining strong buy-in from organizational leaders and frontline staff, (2) delegating responsibility and flexibility for continuing the program to participating sites, and (3) aligning their programs with state or federal initiatives. Lacking sufficient funding, nine awardees had to end their programs.

Although most awardees developed a payment model they pursued during their award, they struggled to implement these models in full, and typically had to obtain internal or other external sources of funding to sustain their programs. The most common payment arrangements implemented were fee-for-service (FFS) billing codes, followed by per beneficiary per month (PBPM) fees and shared savings arrangements through accountable care organizations. Awardee efforts to implement more advanced value-based payment arrangements that would provide flexibility to sustain program services proved more difficult to achieve. Awardees had to pursue arrangements with payers before having sufficient data on whether their programs would achieve their intended effects on service use and costs, so were unable to demonstrate that their programs would save payers money or finalize payment rates. Still, several awardees remained hopeful about gaining this evidence in the future to continue their negotiations and eventually gain payer interest.

Potential lessons for testing future innovations

The evaluation of the HCIA R2 awardees highlighted nine lessons with implications for implementing programs to promote innovation in health care delivery and payment models in the future.

1. Even though there was considerable investment in the innovations, there is little evidence showing the potential for savings to offset the outlay of federal resources under HCIA R2.

2. Although the purpose of the cooperative agreement was to support and facilitate investigator-defined service delivery and payment reform models, the cooperative agreement approach does not lend itself to evaluating the impact of implementation of changes in payment policy.
3. Despite perceived value among clinicians, convincing payers and plans to cover the cost of innovative programs is difficult within a fee-for-service environment or without clearer evidence that savings to the payer are likely to exceed the cost.
4. It is very difficult for innovative programs to move the needle on health care delivery reform in three years.
5. The impact findings suggest that it is easier to reduce ED visits than it is to reduce hospitalizations. However, reductions in ED visits are unlikely to achieve the same level of cost savings as reductions in hospitalizations.
6. Program effects tend to be concentrated among subgroups of higher-risk Medicare and Medicaid beneficiaries most likely to benefit from enhanced intervention services. Future tests of programs might require awardees to specify and collect IDs on one or two subgroups for whom they expect their intervention to be especially powerful.
7. Even when programs achieve favorable results, it is difficult to produce large enough savings from health care delivery innovations to cover or exceed the cost of the intervention, especially within a limited implementation period.
8. Very different types of programs can achieve favorable results, suggesting there could be many opportunities to deliver better care. The findings from this study suggest there are a variety of health conditions and patient populations for which programs can find ways to improve care, and various ways to accomplish it.
9. Programs with evidence of favorable effects shared a few characteristics that were less evident among programs without evidence of favorable effects. These include having previous experience with similar programs and implementing a patient-focused intervention targeted at a socially fragile population. They also include having behavioral health, telehealth, and health information technology as components of the intervention and engaging nonclinical frontline staff.

Conclusions

The HCIA R2 funded a diverse set of awardees to implement innovative solutions for reducing health care service use and costs while improving quality, and to propose payment models to support those innovations. Similar to the findings of the HCIA R1 evaluation,[3] there were mixed findings for HCIA R2: only two programs generated statistically significant estimates of savings. (Two other programs generated statistically significant estimated reductions in service use, but Medicaid data to evaluate the impact on spending were not available). Further, it was not possible to evaluate the impact of 19 programs due to selection bias and small samples. Despite

[3] https://downloads.cms.gov/files/cmmi/hcia-metaanalysisthirdannualrpt.pdf.

considerable investment in HCIA R2, there is little evidence showing potential for savings to offset the outlay of federal resources. Although the purpose of the cooperative agreement was to support and facilitate investigator-defined delivery system and payment reform models, the cooperative agreement approach to identify promising models does not lend itself to evaluating the impact of implementation of changes in payment policy.[1]

Although the study identified only four programs as having evidence of favorable effects on health care costs, hospitalizations, or ED visits during the study period, many awardees continued their programs in whole or in part after the awards ended, demonstrating that organizations found value in these programs. The four programs with evidence of favorable effects shared several characteristics that were less common among programs without evidence of favorable effects. All four had prior experience implementing their intervention, and all served a socially fragile population. Five other characteristics were also associated with evidence of favorable effects, including incorporating a behavioral health component in the intervention, using telehealth services or health IT, relying on nonclinical frontline staff (though only among awardees serving socially fragile people), and having a patient-focused intervention. These findings suggest that CMS or other payers seeking to reduce health care expenditures and improve patients' lives within a relatively short period might wish to pay particular attention to programs with several of these features.

This page has been left blank for double-sided copying.

I. INTRODUCTION

A. Background and purpose of the initiative

The Patient Protection and Affordable Care Act of 2010 established the Center for Medicare & Medicaid Innovation (CMMI) within the Centers for Medicare & Medicaid Service (CMS) to test delivery system and payment innovations and encourage the adoption of those that prove successful by creating new national provider payment policies. To receive broad input from the health care field on innovative solutions, CMMI implemented the Health Care Innovation Awards (HCIA) initiative. The initiative made awards to organizations that enabled them to build on their own experiences to test whether refining and broadening current approaches or implementing new ones yield the desired improvements and efficiencies in delivering health care.

There have been two rounds of funding under the HCIA initiative, with significant differences between them. During the first round (HCIA R1), which began in July 2012, CMMI awarded 108 three-year cooperative agreements to organizations from all 50 states, plus the District of Columbia and Puerto Rico. The purpose of the first round of awards was to test a wide range of service delivery models with the potential to improve outcomes for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries. The second round of funding (HCIA R2) built on the first round but differed in two important ways. First, HCIA R2 focused on four key areas: (1) reducing Medicare, Medicaid, or CHIP costs in outpatient or post-acute care settings; (2) improving care for patients with special needs; (3) testing new financial and clinical models for specific provider types; and (4) improving the health of specific populations by enhancing patient engagement, disease prevention, and wellness efforts. Second, CMS specifically sought new payment models to support the funded service delivery innovations.

Under HCIA R2, CMMI awarded 39 cooperative agreements to organizations from 27 states and the District of Columbia. However, two years after receiving its award, one organization terminated its agreement with CMMI and withdrew from HCIA R2, leaving 38 awardees to complete the funding cycle. Funding began in September 2014 and lasted for an initial three-year period. CMMI granted no-cost extensions for up to 12 months to 30 awardees, enabling them to complete ongoing episodes of care, enroll and serve additional participants, or close out administrative or evaluation activities.

B. Goals of the HCIA R2 evaluation and purpose of this report

CMMI evaluated each of the 38 awards funded under HCIA R2 and described awardees' experiences with developing and implementing alternative payment models. The evaluation had six objectives (text box). The first three annual evaluation reports focused primarily on evaluating the effectiveness of program implementation (Objectives 1, 2, and 5) based on

qualitative analyses conducted during the contract period.[4] This final evaluation report builds on prior evaluation reports by assessing Objectives 3, 4, and 6. The report presents the results from the evaluation of program impacts on health care costs, service use, and quality of care. Based on the findings from the implementation and impact evaluations, the report also identifies program features associated with demonstrated or promising effects on key outcomes.

The HCIA R2 evaluation has six objectives.

1. Describe the implementation experience of each awardee and assess the barriers and facilitators associated with the awardee's success in promoting change.
2. Assess for each awardee the experience of participants, the attitudes of clinical and nonclinical staff toward the model and their work, and participant and staff perceptions of the intervention's effects on the processes and outcomes of care.
3. Assess the effects of each model on health care costs, service use, and quality of care by using the same methodologies and outcome measures across awardees when possible, plus additional outcome measures tailored to each program as appropriate.
4. Synthesize the findings from the implementation and impact evaluations of each program to identify model components and program contexts that appear to be most critical to success and to provide evidence to CMMI about program sustainability, scalability, and replicability.
5. Describe the awardees' payment model designs and their experience in developing and testing the models.
6. Conduct a meta-evaluation of the awardee-specific results, searching for program features consistently associated with successful implementation and program impacts.

C. Challenges to finding global policy lessons across diverse awardees

Although HCIA R2 enabled CMS to test a wide range of novel health care delivery and payment models simultaneously, it also introduced a unique set of evaluation challenges. First, the diversity of the HCIA R2 awards made it challenging to identify global policy lessons about best practices for reducing health care spending and improving quality. As described in Chapter II, the 38 programs differed on several key dimensions, including the target population, intervention services provided, expected outcomes, and program setting. As a result, it was difficult to determine which of those features most directly corresponded to achievement of cost savings.

Second, the cooperative agreements encouraged awardees to continue to innovate and institute changes to their program designs. In response, many programs updated features of their enrollment criteria, intervention services, and staffing roles and responsibilities, among other things. These changes sometimes resulted in a lack of a clear, standardized intervention protocol and made it challenging to determine the definition and scope of the intervention being evaluated, and the timing of expected changes in program impacts over time.

[4] Previous annual HCIA R2 evaluation reports are available at https://innovation.cms.gov/initiatives/Health-Care-Innovation-Awards/Round-2.html.

Third, as discussed in Chapter II, the ability to estimate impacts for some awardees was seriously impeded by the ways they recruited and enrolled participants into the program, the number of enrolled beneficiaries who could be identified in health care claims data, and the availability of data on key outcomes. As a result, it was possible to evaluate the impact of interventions on health care costs or service use for only half (19) of the 38 awardees. For the remaining 19 awardees, it was not possible to provide credible estimates of the program's impact on health care costs and service use for several reasons: it was not possible to identify a credible comparison group, the sample size was not large enough to support an evaluation, or useable data on program outcomes were not available. As a result, the analyses for these awardees only describe the characteristics of program participants at enrollment and program implementation experiences.

D. Road map to the report

This is the first part of a two-part report. This part (**Part I**) summarizes and synthesizes the main impact findings of the individual awardee evaluations. **Part II** includes the 38 awardee-specific evaluation reports, which formed the basis for the synthesis.

The remainder of this part of the report presents the following:

- **Chapter II** describes the criteria used to determine the feasibility of producing estimates of program impacts on expenditures and service use, and the characteristics of the programs that did and did not meet those criteria.
- **Chapter III** summarizes findings from the evaluations of the 19 awardees for which program impacts on expenditures and service use were estimated. It also highlights features that distinguish programs with evidence of favorable program effects from the programs for which the evidence suggests that such effects are unlikely.
- **Chapter IV** describes awardees' efforts to sustain their programs after the end of HCIA R2 funding, and their efforts to implement their proposed payment models to cover the cost of intervention services.
- **Chapter V** discusses the implications of the overall findings from the evaluation and future directions for improving health care service delivery and payment.

The appendices present the methodological details. **Appendix A** provides technical details on the various regression models used to estimate program impacts, the results of which appear in the body of this report. **Appendix B** provides technical details on the propensity score matching approach used to identify external comparison groups for evaluating program impacts on expenditures, service use, and quality. **Appendix C** provides information on the minimum detectable effects on expenditures and service use for each of the program impact evaluations. **Appendix D** provides technical details on the Bayesian analytic methods, which express impact estimates in terms of their probability of being achieved. **Appendix E** describes the features of the models used to evaluate program impacts on expenditures and service use. **Appendix F**

provides a justification for the criteria used to assess whether programs had evidence of favorable impacts on expenditures and service use. **Appendix G** provides a list of variables used to assess the association between awardee characteristics and program impacts. **Appendix H** summarizes the sustainability status and funding source for each awardee's program, including a description of their payment models.

II. IDENTIFYING AWARDEES FOR WHICH CREDIBLE IMPACTS COULD BE ESTIMATED

This chapter describes the criteria used to determine the feasibility of producing estimates of program impacts on health care costs, service use, and quality for each of the 38 awardees, and it identifies the program evaluations that met those criteria. The chapter then identifies program characteristics that were more (and less) common among programs with impact estimates compared to those for which it was not possible to produce impact estimates.

A. Criteria for producing program impact estimates

A key objective of the HCIA R2 evaluation was to assess the effects of each program intervention on health care costs, service use, and quality of care. However, as described in the first three annual evaluation reports, the 38 programs funded under HCIA R2 varied on multiple dimensions, including intervention services provided, target population, where the program was implemented, number of participants, and expected effects on outcomes. These differences created multiple challenges for measuring program impacts.

To produce estimates of program impact, each awardee evaluation had to meet three criteria:

1. A credible comparison group had to be identifiable using claims data.
2. The number of participants had to be large enough to detect a 20 percent effect of the program on total expenditures, number of hospitalizations, or number of emergency department (ED) visits.
3. The key outcomes that the program expected to affect had to be measurable using Medicare or Medicaid claims.

Of the 38 program evaluations, 19 met these criteria, making it possible to estimate intervention impacts on health care cost and service use (Table II.1). Two of these evaluations were based on randomized controlled trials (RCTs). For the other 17 programs, it was possible to identify a comparison group that closely matched the characteristics of the program participants (or the pool of eligibles from which they were drawn) and that minimized the risk of selection bias. It is not feasible to rule out the possibility that selection bias or other confounding factors influenced estimates of program impact given the nonexperimental design of these evaluations. However, the evaluations used the pool of eligibles as the treatment group to avoid selection bias when it seemed likely, and found no strong evidence that confounding factors unrelated to the interventions influenced the main findings. Nine of the evaluations with impact estimates were based on Medicare beneficiaries only, 6 were based on Medicaid beneficiaries only, and 4 included both Medicare and Medicaid beneficiaries.

Table II.1. List of awardees with estimates of program impacts, by payer

Payer	Awardee	Number of awardees
Medicare only	AAMC, Avera, CCC, CCNC, CHIIC, FSCL, NMC, UCSF,[a] UKS	9
Medicaid only	Altarum, Amerigroup, Montefiore, NACHRI, SCH, UIC[a]	6
Medicare and Medicaid	FPHNY, Mesa, NYC H+H, VCC	4

[a] Program impacts based on randomized controlled trial.

B. Programs without impact estimates

For the other 19 awardees, there were serious concerns about conducting a rigorous impact analysis based on one or more of the three criteria. Table II.2 lists the programs without impact estimates, along with the evaluation criteria that were not met. Of these programs, 10 served primarily Medicare beneficiaries, 8 served primarily Medicaid enrollees, and 1 served Medicare and Medicaid beneficiaries.

In all 19 cases, there were insurmountable challenges to identifying a credible comparison group using claims data. Program features that made it impossible to identify a credible comparison group included program eligibility criteria that were not clear or not consistently applied, or those based on clinicians' judgment rather than the occurrence of a specific medical event or diagnosis identifiable in claims data. Selective recruitment based on clinician judgement (program selection) and enrollment being limited to eligible beneficiaries who chose to participate (self-selection) raised concerns about selection bias, which could be either adverse or favorable. The risk of selection bias made it impossible to estimate intervention impacts for program participants. It was sometimes possible to avoid the risk of selection bias by estimating the intervention impact for the entire eligible population (regardless of whether they received intervention services), instead of only for program participants. However, this was only feasible for programs with a participation rate of 15 percent of the eligible population that served a large number of individuals. When the participation rate was less than 15 percent, estimating program impacts for the entire eligible population would have been too imprecise to provide reliable inferences about program effectiveness.

Another common reason for being unable to produce a credible impact estimate was having a sample size of participants that was too small to detect a 20 percent effect of the program on total expenditures, number of hospitalizations, or number of ED visits. In most cases (though not all), this meant the awardee enrolled fewer than 500 beneficiaries who met the eligibility criteria assessable with claims data. Some of these programs had a small eligible population to begin with, whereas others served only a small portion of their eligible population as identified in claims data either because they could not reach all eligible individuals or because those individuals declined to participate. In other cases, a substantial portion of program participants did not appear to meet the awardee's stated eligibility criteria according to claims data.

For example, if a program's eligibility criteria required an ED visit but a patient did not have a claim for an ED visit near the enrollment date, that individual was not identified as a participant even if the awardee reported enrolling the individual in the program. Participants who did not meet the awardee's eligibility criteria according to claims data were not included in the impact analysis because it was not possible to identify comparable comparison cases for them.

Finally, in four cases, it was not possible to conduct a credible impact evaluation because the evaluation could not measure any program outcomes the awardee expected to affect using claims data. For example, Columbia's MySmileBuddy program aimed to reduce early childhood caries. Evaluating the program using claims data was challenging in part because hospitals and other providers frequently do not file dental procedure claims with Medicaid, meaning claims data might not reliably reflect the changes in the number of oral health-related procedures performed.

Table II.2. Awardee evaluations without estimates of program impacts, and the reasons for being unable to produce them

	Reason for being unable to produce credible impact estimates		
Awardee	Inability to identify a credible comparison group	Too few participants to detect meaningful effects[a]	Unable to measure key program outcomes from available claims data[b]
Medicare (N = 10)			
ACCF	X		
Hopkins	X	X	
Icahn	X	X	
Northwell	X	X	X
UHCMC	X		
UMich	X	X	
UNC	X	X	
UNM	X		
Ventura	X	X	
Yale	X		
Medicaid (N = 8)			
BMC	X	X	
CHS	X	X	X
Clifford Beers	X	X	
Columbia	X	X	X
DMC	X		
NHCHC	X		
Wash U	X		X
WI DHS	X		

Table II.2 (*continued*)

Awardee	Reason for being unable to produce credible impact estimates		
	Inability to identify a credible comparison group	Too few participants to detect meaningful effects[a]	Unable to measure key program outcomes from available claims data[b]
Medicare and Medicaid (N = 1)			
UCSD	X		

Note: Payer categories are based on the primary target population served by the awardee's program.

[a] This occurred either if the program did not enroll a sufficient number of eligible individuals or if a large number of beneficiaries enrolled in the program did not appear to meet the program's eligibility criteria in claims data.

[b] This occurred if it was not possible to measure the key outcome or outcomes that the program expected to affect using available claims data.

For 15 of the 19 programs for which a rigorous impact analysis was not feasible, it was at least possible to identify program participants and their core outcomes using Medicare or Medicaid claims data (Table II.3). As a result, it was possible to attempt to understand the program's impact on the health care expenditures, service use, and quality of care using a range of evaluation designs, including difference-in-differences and cross-sectional designs as described in Chapter III, and by working to identify a valid comparison group. Nevertheless, in all cases, there was evidence that the estimates of program effects would have remained biased and, therefore, would have failed to represent a credible estimate of the true impact of the program despite these efforts. Furthermore, it was not possible to estimate intervention impacts over the entire eligible beneficiary population for these awardees because their low participation rates (for example, Yale's participation rate among eligible ED participants was about 3 percent). For the remaining 4 programs, it was not possible to identify participants or their outcomes in administrative data and, therefore, it was not possible to attempt to evaluate the impact of these interventions.

Table II.3. Ability to identify participants in claims data for programs without impact estimates

Was it possible to identify participants in claims data?	Medicare	Medicaid	Number of awardees
Yes	ACCF, Icahn, JHU, Northwell, UHCMC, UMich, UNC, UNM, Ventura, Yale	Clifford Beers, Columbia, DMC, UCSD, WI DHS	15
No	NHCHC	BMC, CHS, Wash U	4

For the 15 programs for which it was possible to identify program participants using claims data, the evaluation generated descriptive information about baseline demographic characteristics, health status, and service use and expenditure data for at least a subset of the program participants (see individual awardee reports). Demographic information typically included the average age of participants and the gender and racial and ethnic distributions. Health status information included factors such as participants' average risk score (expected Medicare or

Medicaid expenditures in the coming year relative to the average for all beneficiaries, based on their chronic conditions), and other program-relevant health factors, such as the proportion on both Medicare and Medicaid. Baseline service use outcomes and expenditures included the number of hospitalizations and ED visits per 1,000 participants and average total Medicare or Medicaid expenditures. For the remaining 4 programs, there was little information about the baseline characteristics of program participants and the awardees generally provided any information that was available.

C. Characteristics that were more and less common in programs for which credible impact estimates could be produced

Programs for which credible impact estimates could be produced were more likely to share some features and less likely to share others, compared with programs for which valid impact estimates could not be produced (Table II.4). The four interrelated characteristics that programs for which credible impact estimates could be produced were *more* likely to have include (1) using passive enrollment procedures, (2) enrolling participants indirectly, (3) implementing an intervention at the provider or organization level, and (4) having telehealth or health information technology and decision support as key intervention components. Programs that used passive enrollment served everyone meeting the eligibility criteria who sought care at a participating site and were therefore more likely to achieve statistical power and less susceptible to selection bias. Indirect participation refers to intervention components implemented at the organization level—such as implementing staffing changes, trainings, or new technologies—that potentially benefit all eligible individuals but that do not directly provide them with services. Programs with indirect participants as well as programs implemented at the provider or organizational level are also more likely to enroll enough participants to detect meaningful effects and less susceptible to selection bias. Programs with telehealth innovations or with health information technology and decision support innovations were also more likely to be implemented at the provider or organization level. The fact that these types of program innovations are well represented in the 19 impact evaluations is a strength as telehealth and health information technology can help to overcome primary care physician shortages, improve patient access to care, and address communication issues in the future.

The characteristics that programs with impact estimates were *less* likely to share than those without impact estimates include (1) targeting acute medical conditions or primary and preventive care, (2) using active enrollment procedures, (3) implementing an intervention aimed at increasing patient engagement or extending provider roles, and (4) serving urban or suburban areas only. Programs targeting beneficiaries with acute or subacute conditions and programs targeting primary and preventive care were likely unrepresented in the impact evaluations because it is generally more difficult to identify beneficiaries who are eligible for these programs within claims data than it is to identify beneficiaries with chronic or complex conditions. Programs that relied on active enrollment (that is, those that required informed patient consent or some other form of direct contact with eligible individuals) had more difficulty recruiting and

engaging patients and were more susceptible to selection bias. Beneficiaries who were referred into or chose to participate in these programs potentially differed in important unobserved ways from beneficiaries who were not referred or decided not to participate. Most programs with interventions that aimed to increase patient engagement or extend provider roles used active enrollment and were therefore also more susceptible to selection bias.

Table II.4. Characteristics of awardees with versus without credible impact estimates

Program characteristic	Total number of awardees	Number of awardees with credible impact estimates	Percentage of awardees with credible impact estimates
Total awardees	38	19	50%
Target payer[a]			
Medicare	19	9	47%
Medicaid	14	6	42%
Both	5	4	80%
Target population			
Youth with complex conditions	5	3	60%
Adults with chronic conditions—low risk	8	4	50%
Adults with chronic conditions—high risk[b]	6	4	67%
Behavioral health and cognitive disorders	5	3	60%
Acute and subacute conditions	6	2	33%
Primary and preventive care	8	3	38%
Type of enrollment[c]			
Active	24	7	29%
Passive	11	10	91%
Both	3	2	67%
Type of participant[d]			
Direct	20	10	50%
Indirect	8	6	75%
Both	10	3	30%
Level of intervention[e]			
Patient level	36	17	47%
Provider or organization level	11	8	73%
Key innovation[e]			
Care coordination	20	8	40%
Patient engagement and education	10	3	30%
Care transitions	10	6	60%
Health IT and decision support	11	9	82%

Table II.4 (*continued*)

Program characteristic	Total number of awardees	Number of awardees with credible impact estimates	Percentage of awardees with credible impact estimates
Telehealth	5	4	80%
Extended provider roles	6	2	33%
Program setting[e]			
Provider based	28	14	50%
Community based	14	8	57%
Home based	12	6	50%
Virtual	8	7	88%
Market area			
Served a rural area	14	10	53%
Served urban or suburban areas only	24	9	38%
Number of individuals served			
500 or fewer	2	0	0
501 to 1,000	6	3	50%
More than 1,000	30	16	53%

[a] CMMI required that programs target Medicare, Medicaid, and/or Children's Health Insurance Plan (CHIP) beneficiaries. Although some programs targeted individuals directly based on their insurance status, others enrolled individuals regardless of their coverage, meaning participants may have included individuals with public and/or private insurance. Payer type represents the primary target population of the intervention. For five programs, eligible individuals included a sizable portion of both Medicare and Medicaid beneficiaries. It was feasible to produce credible impact estimates for four of these five programs and in each case, the evaluation produced estimates of program impact for both beneficiary types.

[b] High risk is defined as having a precipitating inpatient or ED service that triggers enrollment into the program or having clinical conditions associated with a high risk of having inpatient or ED service use in the coming year.

[c] Some programs required informed consent or some other form of direct contact with potential participants triggered by a predefined event (active enrollment), whereas others served everyone who sought care at a participating site who met the program eligibility criteria (passive enrollment). A few programs had multiple components, some of which involved passive enrollment and others that involved active enrollment.

[d] A direct participant is an individual who receives care or services paid for by HCIA R2 program funding, such as care coordination services. An indirect participant is anyone who does not receive such services, but who benefits from the HCIA R2 funding nonetheless. For these participants, HCIA R2 funding is generally used to assist service providers, such as funding to hire program staff, train intervention staff, and purchase or develop technology. These resources, in turn, can enhance and support clinicians' ability to deliver high quality, cost-efficient care to participants. A few programs had multiple components, some of which required direct participation and some for which participation was indirect. For example, CHIIC included both care coordination services, which required direct participation, and organizational-level quality improvement activities. The benefit to participants resulting from these activities was indirect.

[e] Categories are not mutually exclusive, and therefore rows will not sum to 38 programs (100 percent).

CHIIC = Catholic Health Initiatives Iowa Corp.; ED = emergency department; HCIA R2 = Round 2 of the Health Care Innovation Awards; IT = information technology.

This page has been left blank for double-sided copying.

III. SUMMARY OF FINDINGS FROM IMPACT ANALYSIS

This chapter describes the results of the impact evaluations designed to assess the effects of each intervention on health care costs, service use, and quality of care. It also identifies the programs that had evidence of favorable impacts on those core outcomes and describes the program characteristics, implementation experiences, and other factors that distinguish programs with evidence of favorable effects from those without evidence of favorable effects.

A. Impact evaluation design overview

As described in Chapter II, there were 19 programs for which it was possible to estimate the impact of the intervention on core outcomes (text box; Table III.1). Given the diversity among these programs, accurate evaluation of their impacts required a variety of designs and analytic approaches. This section describes the evaluation designs, analytic models, and beneficiary populations used to assess program impacts.

Program impact evaluations included four core outcomes, when feasible

- Total Medicare or Medicaid expenditures ($ per beneficiary per month)
- Number of hospital admissions (per 1,000 beneficiaries)
- Number of ED visits (per 1,000 beneficiaries)
- Proportion of beneficiaries with a 30-day hospital readmission

For 2 of the 19 programs with impact estimates, the evaluation used RCTs to estimate intervention impacts. The other 17 evaluations compared outcomes for program participants or individuals eligible for the program to those of a matched comparison group. **Appendix B** provides additional technical details on the matching approach used to identify comparison groups.

Program evaluations used one of two analytic models: difference-in-differences (13 programs) or cross-sectional (6 programs). Difference-in-differences models assume that in the absence of the intervention, the change in outcomes between the pre-program and program periods would have been the same, on average, for the treatment group as for the comparison group. Therefore, any difference in the change in outcomes should be attributable to the intervention. For most of the difference-in-differences models (12 of 13 programs), the baseline and follow-up period observations were on the same individuals. However, for one awardee (AAMC) that implemented a provider-level intervention affecting all eligible beneficiaries treated at participating facilities, the baseline observations were on the cohort of eligible individuals treated at the same facilities in the year before program start-up and the follow-up period observations were on the cohort of eligible individuals treated after program launch.

The evaluation used post-period cross-sectional models for the programs evaluated with an RCT (two programs), as well as when baseline outcomes occurred before a major shift in a

participant's health trajectory, meaning that the change over time in a beneficiary's outcomes was not a relevant indicator of the program's impact (four programs). This occurred when, for example, program enrollment required a precipitating event and, therefore, baseline values of outcomes were quite different from follow-up period values. **Appendix A** provides additional technical details on the regression models (including difference-in-differences and cross-sectional designs) used to estimate program impacts on core outcomes.

For most (15) of the evaluated programs, analytic models were estimated over intervention participants only, regardless of level of engagement; however, for the other 4 programs, the treatment group comprised the entire population of individuals who were eligible to participate, according to claims data. Individuals were classified as program participants if they formally enrolled in the program or if they directly or indirectly received any of the intervention services provided by the program, regardless of whether they actively engaged in the program or received all available services. Of the 15 programs, 6 (AAMC, Altarum, Avera, CCC, NACHRI, and UKS) implemented their intervention, or a component of it, at the provider level and used passive enrollment, meaning the program served everyone who sought care at a participating site who met the program eligibility criteria. For these six programs, the participant population and the eligible population were the same.

For the four programs using all eligibles as the treatment group (CCNC, FSCL, Mesa, and NYC H+H), the treatment group included beneficiaries if the impact evaluation identified them as being eligible for the program using Medicare or Medicaid claims data, even if those beneficiaries chose not to participate or did not receive any intervention services. The impact evaluations used this broader treatment-eligible population when it seemed likely there was a high risk that adverse or favorable selection into the awardee's program would bias the estimated impact on core outcomes. The evaluation estimated impacts over the eligible beneficiary population only when programs had a relatively large number of participants (for example, 700 or more) and a participation rate of 15 percent or greater of the total eligible beneficiary population. If a program had a participation rate of less than 15 percent of the eligible population, estimates of program impact on the core outcomes based on the entire eligible population would have been too imprecise to support reliable inferences about impacts unless the number of enrollees was extremely large. For example, under some realistic assumptions, as many as 37,400 eligible beneficiaries, of whom 3,740 were participants, would be needed to detect a 20 percent effect on the participants, if the participation rate was only 10 percent. However, if the participation rate was 30 percent, this level of precision would require only about 1,250 participants (4,167 total eligible beneficiaries).

As described in Chapter II, to produce credible impact estimates, evaluations needed to have a large enough sample size to detect program effects less than or equal to 20 percent on total expenditures, number of hospitalizations, or number of ED visits. **Appendix C** provides details on the minimum effect that each of the 19 impact evaluations was able to detect for each of expenditure and service use outcomes. It is important to note that the minimum detectable effect had to be estimated in advance of a full data analysis, based on assumptions about the expected

number of participants who could be identified in claims data and the variability in outcomes among these participants. In several cases, the usable treatment group sample was smaller and their outcomes more variable than the original assumptions projected. As a result, there were four awardees for which the impact evaluation was not able to detect (with 80 percent confidence) impacts of 20 percent on any of the core outcomes (Amerigroup, Montefiore, SCH, and UCSF).

In addition to the analytic models described earlier, all 19 programs were evaluated using a Bayesian analytic approach. This approach supplements the main analysis by drawing on estimates from the HCIA R1 studies to sharpen the estimates of impacts on core outcomes, and by estimating the likelihood that the true impact falls in certain policy-relevant ranges. **Appendix D** reports the technical details of the Bayesian analyses and high-level results. The results largely corroborate the findings presented in this chapter.

Table III.1. Evaluation design features for awardees with impact estimates

Awardee	Number of treatment beneficiaries in impact estimate	Participation rate	Evaluation design	Impact estimate based on participants or eligible beneficiaries	Study sample used to estimate impact	Post-enrollment follow-up period used to estimate impact
Medicare (N = 9)[a]						
AAMC	145,938	100%	DD[b]	Participants[c]	Full sample	1–3 months
Avera	7,194	100%	DD	Participants[c]	Long-term NF residents	1–24 months
CCC	900	Unknown	CS	Participants[c]	Full sample	1–12 months
CCNC	110,968	55%	DD	Eligibles	Full sample	1–24 months
CHIIC	1,447	23%	DD	Participants	Full sample	13–24 months
FSCL	2,097	38%	CS	Eligibles	Full sample	1–12 months
NMC[d]	430	8%	CS	Participants	Full sample	1–12 months
UCSF	358	Unknown	RCT	Participants	Full sample	1–12 months
UKS	702	100%	CS	Participants[c]	Patients with AMI treated at CAHs	1–12 months
Medicaid (N = 8)[a]						
Altarum	94,944	100%	DD	Participants[c]	Full sample	1–24 months
Amerigroup	299	30%	DD	Participants	Full sample	1–12 months
Mesa[e]	2,872	75%	DD	Eligibles	Full sample	1–12 months
Montefiore	1,758	Unknown	DD	Participants	Full sample	13–24 months
NACHRI	3,528	100%	DD	Participants[c]	Full sample	13–24 months

Table III.1 (*continued*)

Awardee	Number of treatment beneficiaries in impact estimate	Participation rate	Evaluation design	Impact estimate based on participants or eligible beneficiaries	Study sample used to estimate impact	Post-enrollment follow-up period used to estimate impact
NYC H+H[e]	9,747	15%	DD	Eligibles	Medicaid enrollees eligible during first 9 months of program	1–12 months
SCH	516	17%	DD	Participants	Full sample	13–24 months
UIC	821	20%	RCT	Participants	Higher risk children	1–24 months
Medicare and Medicaid (N = 2)[a]						
FPHNY[d]	1,637	3%	DD	Participants	Full sample	1–36 months
VCC[d]	2,952	2%	DD	Participants	Full sample	1–36 months

[a] The program evaluations identified the public payer (Medicare or Medicaid) providing insurance coverage to a majority of individuals eligible for the intervention. If programs included a sizable portion of both Medicare and Medicaid beneficiaries, impacts were estimated for both types of beneficiaries.

[b] Impact evaluation relied on a repeat cross-sectional difference-in-differences design. All other difference-in-differences models relied on longitudinal designs.

[c] These awardees each featured passive enrollment, meaning that the intervention included all eligible individuals because it was at the facility level, and patient consent was not required. Thus, the participation rate is 100 percent by design.

[d] The minimum participation rate threshold of 15 percent pertained only to awardees for which impacts were estimated using all eligible beneficiaries. The participation rate was not applicable when impacts were estimated over participants only.

[e] The program enrolled both Medicare beneficiaries and Medicaid enrollees. Assessments of favorable impacts relied on eligible Medicaid enrollees only. For Mesa, this was because Medicaid enrollees accounted for 60 percent of the study sample and the participation rate was higher among this subgroup. For NYC H+H, this was because favorable program effects were limited to Medicaid enrollees eligible for the program during the first nine months.

AMI = acute myocardial infarction; CAH = critical access hospital; CS = cross-sectional; DD = difference-in-differences; NF = nursing facility; RCT = randomized controlled trial.

B. Methods for assessing evidence of favorable impacts on core outcomes

The results of the impact evaluations were assessed to determine whether programs had evidence of favorable impacts on the core outcomes. Developing a common approach to identifying programs that appeared to have a favorable impact was challenging because the evaluations estimated the effects of each program for multiple follow-up periods, subgroups, and outcome measures (Table E.1; **Appendix E**).

Program evaluations generally included multiple 12-month post-enrollment follow-up periods. The number of follow-up periods differed based on data availability and sample size, and to account for variations among programs in expected length of time until intended effects became apparent. In general, interventions addressing acute conditions were expected to have a more immediate effect on core outcomes and, therefore, their effects could be analyzed using a shorter

follow-up period. Interventions addressing chronic or behavioral health conditions were expected to affect core outcomes over the longer term and, therefore, required longer follow-up periods. For example, Mesa expected its Fire and Medical Department's Community Care Response Initiative to have an immediate impact on ED visits because it designed the program to divert low-acuity 911 callers from the ED by dispatching a mobile health unit instead of an ambulance. Given the short-term nature of the intervention, the evaluation included only 12 months of follow-up to avoid unwarranted attenuation of the impact estimates. In contrast, CHIIC's Transitioning a Rural Health Network to Value-Based Care program implemented population health activities, such as health coaching in primary care clinics, to address the needs of individuals with specific chronic conditions, expecting that the impact on health expenditures and service use would accrue slowly over time. Accordingly, the evaluation of this program included 24 months (two 12-month periods) of follow-up.

In addition, some programs expected their interventions to have a greater effect on outcomes for specific beneficiary subgroups at increased risk of needing services. The impact evaluations for those programs included relevant subgroup analyses, when sample sizes permitted. For instance, as described later, Avera's eLongTermCare (eLTC) program enrolled both long-term care nursing facility residents and beneficiaries requiring short-term post-acute skilled nursing services. Because of major differences in the health care needs between these two types of beneficiaries, the evaluation examined the impact of the intervention for them separately. Most impact evaluations also examined whether program effects on the core outcomes differed for those enrolling during the first nine months of implementation compared to impacts for those enrolling during the later months, when the programs had time to mature. Finally, as described in Section D of this chapter, in addition to the core outcomes, several program evaluations included program-specific outcomes hypothesized to be affected by the intervention, such as process measures and clinical outcomes.

To determine which programs had evidence of favorable effects on the core outcomes, a consistent set of rules was applied for all awardees. The rules described in the text box were used to identify the most appropriate post-enrollment follow-up period and study sample for assessing the favorability of program impacts on core outcomes. It is important to understand that although the rules for assessing effectiveness are the same, those rules led to focusing on different outcomes, time periods, and subgroups for different awardees. The goal was to ensure that a program could be considered effective even if it did not have statistically significant findings for all outcomes for all enrollees over the full program period. **Appendix F** provides detailed justifications for the selection of the follow-up period and study sample for each program.

The impact assessment applied three rules to determine which programs had evidence of favorable impacts on health care costs and service use for some time periods, subgroups, or outcomes.

Rule 1. If the results were favorable for the full sample, those results were used to assess program impact. If the results were not favorable for the full sample, the assessment examined subgroups of relevance to the intervention, when available. If the results were not favorable for the full population but favorable for a subgroup, the assessment identified the program intervention as having evidence of favorable impact for that subgroup.

Rule 2. If the evaluation results were similar across each 12-month follow-up period (as defined by the intervention), the impact assessment relied on the cumulative results over all follow-up periods. If the results differed across follow-up periods, the impact assessment considered the results for each individual period. If the program had favorable results for the second 12-month follow-up period or later, it was considered to have a favorable impact. If the results were favorable only during the first 12-month follow-up period, the impact assessment identified it as favorable if the awardee's theory of action expected the intervention to have impacts mainly within enrollees' first year.

Rule 3. Assessments of favorable impacts did not rely solely on the statistical significance of the evaluation results. Instead, the assessments also considered the size and consistency of the results across the health care costs and service use outcomes. If a program had at least one favorable and statistically significant estimate that was consistent with the awardee's theory of action, and the estimates for the other core outcomes were not adverse and large, the assessment identified it as having evidence of a favorable impact.

The assessment of favorable impacts focused on three of the four core outcomes: health care costs, hospital admissions, and ED visits. Estimated program effects on the fourth outcome—30-day readmissions— were evaluated for only 11 of the 19 programs and none of the results were statistically significant. Impacts on 30-day readmissions were not calculated for the other 7 programs for three reasons: (1) the program served primarily Medicaid enrollees and there were concerns about the quality of the available Medicaid claims data, (2) the program did not expect to affect the readmission rate, or (3) the evaluation lacked the statistical power to detect meaningful changes in 30-day readmission rates.

C. Programs with evidence of favorable impacts on core outcomes

After applying the assessment rules described in the previous section, 4 of the 19 programs for which impact evaluations were produced showed evidence of favorable effects on one or more core outcomes for all or a subgroup of beneficiaries for at least one 12-month period, and no evidence of adverse effects: Avera's eLTC program, Montefiore's BHIP, UIC's CHECK program, and NYC H+H's ED care management program (Table III.2 and Figures III.1–III.3). The accompanying text boxes describe each of these programs and the evaluation results. Avera's eLTC program primarily served Medicare beneficiaries; Montefiore's BHIP and UIC's CHECK program mainly served Medicaid enrollees. Although NYC H+H's ED care management program served both Medicare beneficiaries and Medicaid enrollees, the favorable

effects concentrated among the Medicaid enrollees eligible during the first 9 months of the program.

The impact evaluation designs and analytic models for these four programs differed. UIC's CHECK program relied on a RCT, whereas the other three programs required selection of a matched comparison group. This difference led to the evaluations of Avera's eLTC, Montefiore's BHIP, and NYC H+H's ED care management program relying on difference-in-differences models, whereas the evaluation of UIC's CHECK program compared the treatment and control group outcomes at follow-up only, controlling for baseline values of the outcomes.

The approach to defining the beneficiary population included in the evaluations and the follow-up periods assessed also differed. In the case of UIC, the evaluation included all Medicaid enrollees randomly assigned to the treatment or control groups. Avera's eLTC program served all eligible individuals, so the participant group was essentially the same as the eligible group. NYC H+H's program enrollment criteria could not be replicated using claims data because the program selectively enrolled patients based on clinicians' judgment, so the evaluation included all beneficiaries who met the eligibility criteria assessable with claims data. Finally, the evaluation of Montefiore's BHIP used a subset of program participants as the treatment group because the eligibility criteria could not be confirmed in the claims for many enrollees; instead, the awardee provided depression screening data for those considered for the intervention. The evaluation included the participants who met the depression screening criteria and a comparison group composed of individuals treated at nonparticipating Montefiore sites and practices who met the same depression screening criteria.

With respect to the follow-up period, for three of the programs (Avera, Montefiore, and UIC), the impact assessment relied on results for either the entire 24-month follow-up period or for the second-year follow-up period only, to demonstrate the longer-term effects on core outcomes. For NYC H+H, the impact assessment relied on the first-year follow-up period because the intervention was short term and expected to have a more immediate effect on outcomes.

Table III.2. Summary of estimated effects on core outcomes used to determine whether programs had evidence of favorable impacts

	Total expenditures ($ PBPM)		Hospital admissions (per 1,000 beneficiaries)		ED visits (per 1,000 beneficiaries)	
Awardee	Impact estimate	Percentage impact	Impact estimate	Percentage impact	Impact estimate	Percentage impact
Medicare (N = 9)[a]						
AAMC	-$10[b]	< -1%	-1[b]	< -1%	3[b]	< 1%
Avera	-$73[b]*	-4%	-10[b]	-1%	-73[b]***	-9%
CCC	$164	6%	13	2%	-212***	-21%
CCNC	$10	< 1%	< 1	< 1%	-5	< -1%
CHIIC	-$13	-1%	-26	-7%	-50	-7%
FSCL	$601*	10%	NA	NA	NA	NA
NMC	-$186[b]	-11%	-10[b]	-1%	-5[b]	< -1%
UCSF	-$81	-5%	-46	-10%	-63	-9%
UKS	-$109	-4%	32	3%	-40	-2%
Medicaid (N = 8)[a]						
Altarum	NA	NA	NA	NA	5	2%
Amerigroup	NA	NA	-39	-19%	-174	-13%
Mesa[c]	NA	NA	9[b]	< 1%	553[b]***	11%
Montefiore	NA	NA	-56	-18%	-177*	-14%
NACHRI	NA	NA	51*	16%	-120***	-10%
NYC H+H[c]	NA	NA	-57*	-6%	-212[b]**	-7%
SCH	$252	9%	97	34%	51	5%
UIC	-$44**	-21%	-17	-15%	-56	-6%
Medicare and Medicaid (N = 2)[a]						
FPHNY	-$6[d]	< -1%	27	3%	52	4%
VCC	-$166[d]	-8%	-11	-2%	76	6%

[a] The program evaluations identified the public payer (Medicare or Medicaid) providing insurance coverage to a majority of individuals eligible for the intervention. If programs included a sizable portion of both Medicare and Medicaid beneficiaries, impacts were estimated for both types of beneficiaries.

[b] The evaluation top-coded the results at the 98th percentile to minimize the sensitivity of the estimates to a few extreme outliers.

[c] The program enrolled both Medicare beneficiaries and Medicaid enrollees. Assessments of favorable impacts relied on eligible Medicaid enrollees only. For Mesa, this was because Medicaid enrollees accounted for 60 percent of the study sample and the participation rate was higher among this subgroup. For NYC H+H, this was because favorable program effects were limited to Medicaid enrollees eligible for the program during the first 9 months.

[d] Total expenditures data are available only for the Medicare beneficiaries.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; NA = not available; PBPM = per beneficiary per month.

Overall, the impact evaluation results demonstrate that two of the four programs (Avera and UIC) had statistically significant reductions in total expenditures for at least a portion of their target population, although for UIC the size and statistical significance of the estimated impact on total expenditures was sensitive to high-cost outliers. Estimated reductions in hospitalizations and ED visits appear to have driven these reductions in total expenditures for UIC; for Avera the primary driver was estimated reductions in ED visits. It was not possible to evaluate the impact of Montefiore's BHIP or NYC H+H's ED care management program on total expenditures due to the absence of data on Medicaid expenditures for managed care plans. However, Montefiore's program had large and statistically significant estimated reductions in ED visits and large but not statistically significant estimated reductions in hospitalizations. NYC H+H's program had moderate and statistically significant estimated reductions in both ED visits and hospitalizations.

For Avera, NYC H+H, and UIC, the evidence of favorable impacts was limited to subgroups of beneficiaries expected to receive the greatest benefit from the program intervention. As previously noted, Avera's eLTC program enrolled both long-term care nursing facility residents and skilled care patients who required short-term skilled nursing services. The impact evaluation results show that the eLTC program had a favorable impact on the service use and expenditure outcomes for the long-term care beneficiary subgroup. Although the impact evaluation demonstrated that the eLTC program also had statistically significant reductions in ED and observation visits for skilled care beneficiaries, the reductions were not large enough to yield a discernable reduction in total expenditures for this subgroup (data not shown; see awardee report). Smaller effects on estimated expenditures for skilled care beneficiaries could be due to this subgroup having less exposure to the intervention before discharge, or to less potential of the intervention to affect outcomes for the shorter-term nursing home residents. NYC H+H's ED care management program enrolled both Medicare and Medicaid beneficiaries. However, program effects concentrated among Medicaid beneficiaries eligible during the first nine months of the program, likely because early enrollees tended to be sicker and to have greater prior use of hospital and ED care compared to individuals who were eligible for the program after those initial nine months. This is due at least in part to the awardee expanding its eligibility criteria toward the end of the first program year to increase enrollment. Finally, results from the impact evaluation of UIC's CHECK program show that program effects concentrated among the higher-risk Medicaid beneficiaries (about one-quarter of its enrollees), defined as children who had one or more ED visits or one or more inpatient hospitalizations in the year before random assignment. This finding is consistent with the expectation that program effects were likely to concentrate among sicker patients.

Although program impacts for Montefiore's BHIP were evaluated over the full sample of participants, the results suggest that reduced visits among individuals with mild to moderate depression at enrollment drove the statistically significant reductions in ED visits for the full sample. In contrast, estimated reductions in hospitalizations, although not statistically significant, were mainly due to a reduction in admissions among individuals with moderately severe depression at enrollment.

Avera Health's eLTC program

The eLTC program provided to staff and residents in nursing facilities a set of geriatric care and telehealth services aimed at helping residents gain access to timely, resident-centered care, thus reducing unnecessary transfers to EDs and hospitals. The program had an estimated 4 percent reduction in total expenditures, driven by a 9 percent reduction in ED and observation visits, both of which were statistically significant, over the 24-month post-enrollment follow-up period for the 7,194 long-term care beneficiaries eligible for the program. The program also had an estimated 10 percent reduction in ED and observation visits for the 2,414 beneficiaries with short-term rehabilitative stays, which was statistically significant. However, these effects were not large enough to lead to a reduction in total expenditures for this subgroup. Three features contributed to the favorable impact of the program: (1) the awardee's prior experience with telehealth services; (2) providing training and support to nursing facility staff during program implementation, which encouraged the use of the telehealth services; and (3) current value-based payment policies for hospitals, which created incentives to reduce hospital readmissions.

Montefiore Medical Center's BHIP

The BHIP aimed to reduce health care costs and service use and improve care quality by integrating behavioral health services in the primary care setting. The program used data from initial mental health screens conducted with individuals at participating primary care sites, as well as follow-up behavioral health scales, to determine appropriate treatment approaches. During the 13- to 24-month post-enrollment follow-up period, the 1,759 program participants included in the evaluation had 14 percent fewer ED visits than the comparison group, which was statistically significant. Program participants also experienced an estimated 18 percent reduction in hospital admissions compared to the comparison group, although this was not statistically significant. Results for months 1 through 12 after enrollment were smaller and not statistically significant, consistent with expectations that effects would not emerge until after significant exposure to the program. The estimated decreases in service use for program participants can be attributed to (1) improving the rates of depression screening and follow-up; (2) integrating the program into existing workflows within primary care practices; and (3) using health IT to enable providers to monitor beneficiaries' progress and adjust treatment promptly.

NYC H+H's ED care management program

The ED care management program aimed to avoid unnecessary hospital and repeat ED visits, and reduce health care costs, by providing care management and 90-day care coordination to eligible individuals who visited the ED at one of six NYC H+H locations. During the 12-month post-enrollment follow-up period, the program had an estimated 8 percent reduction in ED visits and an estimated 6 percent reduction in hospitalizations among 9,747 Medicaid enrollees eligible for the program during its first nine months after implementation; both estimates were statistically significant. Program effects concentrated among Medicaid enrollees eligible during the first nine months, most likely because program participants tended to be sicker and used the ED more frequently than participants enrolling after those first nine months. This suggests that the intervention might be more effective when provided to patients who have been more intensive recent users of health care services. More targeted enrollment of patients in the future could further enhance the effects of the program.

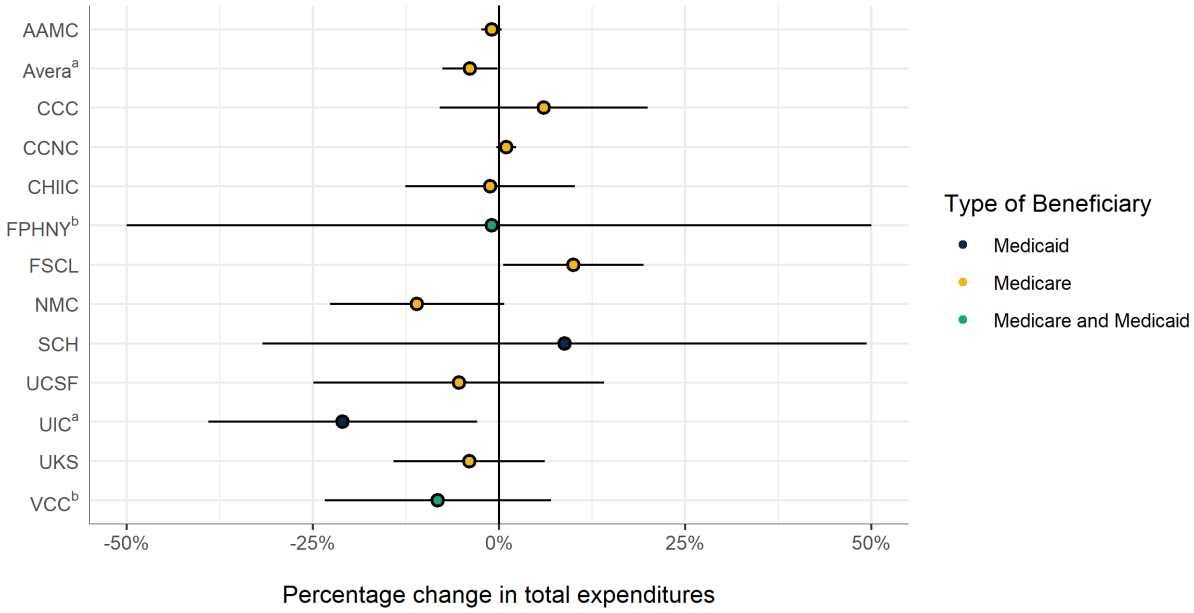
The University of Illinois' CHECK program

The CHECK program aimed to improve health and quality of life and reduce Medicaid costs for children and young adults with asthma, diabetes, sickle cell disease, or prematurity through improved coordination of medical, nonmedical, and mental health services. Among the higher-risk beneficiary subgroup (1,640 children), the CHECK program enrollees had 21 percent lower total expenditures than the control group over the 24-month follow-up period, which was statistically significant. The size of the estimated impact expenditures was somewhat inflated by high-cost outliers in the comparison group. Program enrollees also had an estimated 15 percent fewer hospitalizations and 6 percent fewer ED visits, although neither of these was statistically significant. However, the estimated effects on both the probability of being hospitalized and the probability of having an ED visit were both statistically significant. An important factor that likely contributed to the favorable results is that the RCT to evaluate the program began after a start-up period during which the awardee made several changes (hiring additional case workers and reducing case worker workloads) to resolve implementation challenges.

D. Programs for which the evidence does not suggest favorable impacts on core outcomes

Although the impact assessment found only four programs to have strong evidence of effectiveness, six others had some favorable results but did not meet one of the three criteria provided earlier to be classified as effective. Four programs had consistently favorable results across two or more of the three core outcomes but were not assessed as having evidence of favorable effects: Amerigroup, CHIIC, NMC, and UCSF. For Amerigroup and UCSF, the program evaluations lacked a sufficient sample size to be able to detect reductions of a reasonable magnitude (**Appendix C**). Although the interventions might have had an effect, the estimates were too imprecise to support a favorable assessment. In addition, although Amerigroup had moderate estimated reductions in hospitalizations and ED visits during the first 12-month follow-up period, these results were not statistically significant and were not sustained during the second 12-month follow-up period (data not shown; see awardee report). For CHIIC, although the estimated effects on hospitalizations and ED visits were moderate, the estimated effects on expenditures were small and none of the results were statistically significant. Similarly, for NMC, although estimated effects on total expenditures were moderate, the estimated effects on hospitalizations and ED visits were much too small to have generated expenditure reductions, and again none of the results were statistically significant. Two other programs (CCC and NACHRI) had favorable estimated effects on ED visits, but large, adverse estimates for other core outcomes, which made it untenable to classify these awardees as effective (see Rule 3). **Appendix F** provides a more complete discussion of the rationale for assessing each of these programs as not having evidence of favorable impacts on the core outcomes, based on the rules discussed earlier.

Figure III.1. Summary of estimated percentage impacts (and 90 percent CIs) on total expenditures



Note: For 1 awardee (FPHNY), the CI exceeded ± 50 percent. Six awardees (Altarum, Amerigroup, Mesa, Montefiore, NACHRI, and NYC H+H) of the 19 with impact estimates did not have data available on total Medicaid expenditures.

[a] Indicates awardees with favorable impact estimates.

[b] Total costs of care estimates based on Medicare beneficiaries only.

CI = confidence interval; PBPM = per beneficiary per month.

Figure III.2. Summary of estimated percentage impacts (and 90 percent CIs) on hospital stays

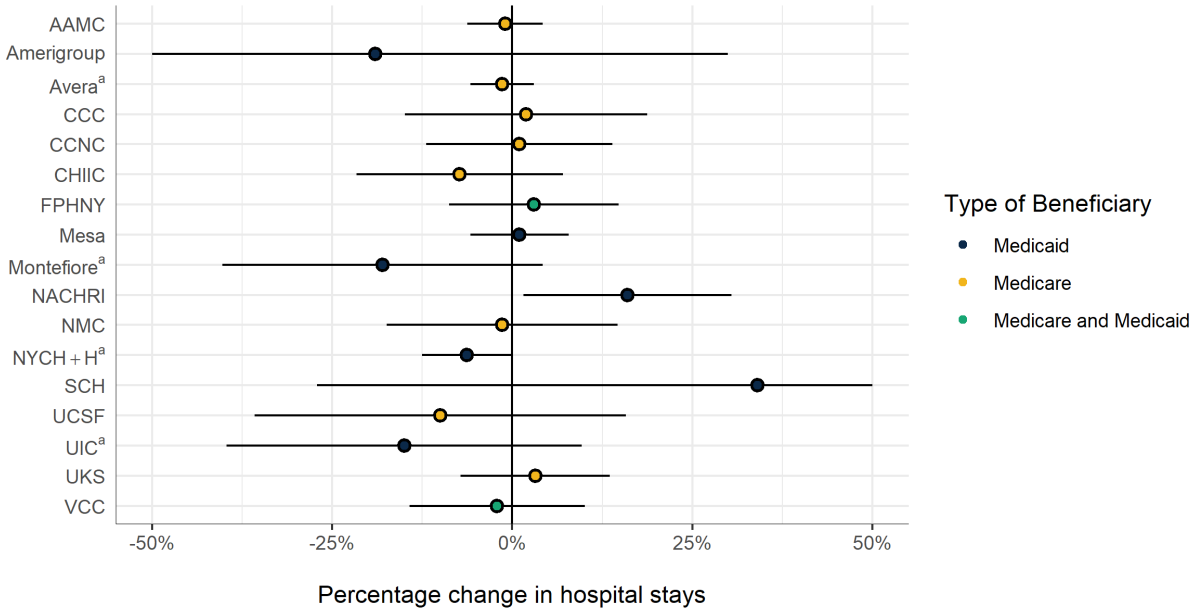


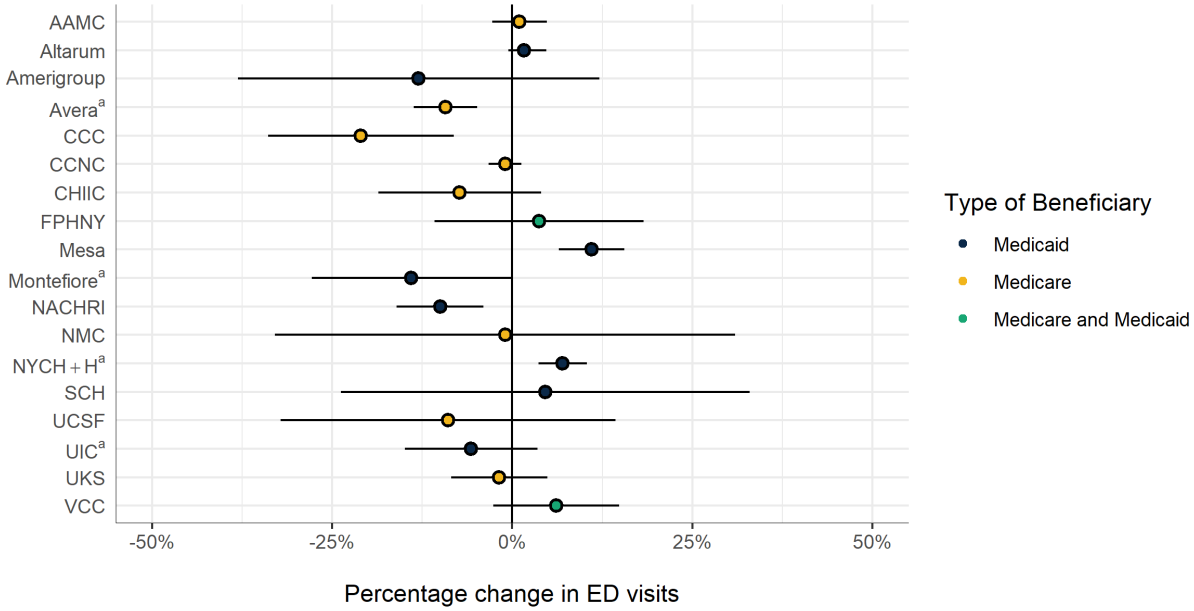
Figure III.2 (*continued*)

Note: For 2 awardees (Amerigroup and SCH), the CIs exceeded ± 50 percent. Additionally, 2 awardees (Altarum and FSCL) of the 19 with impact estimates did not have data available on hospital stays.

[a] Indicates awardees with favorable impact estimates.

CI = confidence interval.

Figure III.3. Summary of estimated percentage impacts (and 90 percent CIs) on ED visits



Note: One awardee (FSCL) of the 19 with impact estimates did not have data available on ED visits.

[a] Indicates awardees with favorable impact estimates.

CI = confidence interval; ED = emergency department.

The remaining programs had impact estimates that were inconsistent across the core outcomes, were highly imprecise, or suggested an increase in service use or total expenditures. Three programs (FSCL, Mesa, and NACHRI) led to statistically significant *increases* in either service use or total expenditures. During the 12-month follow-up period, FSCL led to an estimated increase in total Medicare expenditures among beneficiaries eligible for the program likely because those beneficiaries used more intensive hospice services for a longer time period compared to the control group. Although Mesa's program initially decreased ED visits among eligible Medicaid beneficiaries during the first few months of follow-up, it led to a statistically significant increase in ED visits over the full 12-month follow-up period. This is likely because it is challenging to change behavior of frequent ED users over the longer term; the short-term decreases in ED use were not sustained over time. Finally, although NACHRI's program led to statistically significant reductions in ED visits during the second year of follow-up among Medicaid beneficiaries who participated in the program, it also led to a statistically significant increase in hospitalizations, as previously noted. This may have been because program participants had more complex needs than individuals in the comparison group and as a result, hospitalizations may have been largely unavoidable.

It is important to note that some of the programs did not expect to improve the core outcomes during the one- to three-year follow-up period. As a result, several program evaluations included additional outcomes, measurable using claims data, that the programs intended to improve within the observable time frame. For example, FPHNY's Project INSPIRE aimed to improve initiation of and adherence to hepatitis C treatment and, therefore, was examined for improvements in hepatitis C-specific outcomes. As expected, this evaluation did not find large or statistically significant estimates of reductions in total expenditures in the program period, but the program did have a sizeable and statistically significant estimated increase in hepatitis C drug treatment prescription fills among enrolled beneficiaries. In addition, Altarum's Michigan Caries Prevention Program aimed to improve dental and oral health outcomes for children and reduce associated costs and, therefore, the awardee did not expect the intervention to have a significant impact on total medical expenditures, hospitalizations, or ED visits. Instead, the impact evaluation included receipt of dental and oral health care, and other dental and oral health outcomes. The program's estimated increases in patients' receipt of fluoride varnish, oral health evaluations, and preventive dental visits all were statistically significant. These favorable preventive care effects did not, however, lead to statistically significant estimates of increases in the receipt of dental sealants or decreases in restorative procedures.

E. Features that distinguish programs with versus without evidence of favorable impacts

A key overall goal of the evaluation was to determine if any program features were strongly associated with an awardee having favorable impacts on total expenditures, hospitalizations, or ED visits. This section combines the qualitative findings from the evaluation of program implementation[5] with the quantitative findings from the program impact evaluations to identify program features associated with evidence of favorable effects. The analysis includes only 14 of the 19 programs for which impacts were estimated—the 4 identified in Section C that had favorable impact estimates and 10 that did not.[6]

The analysis of program features associated with favorable outcomes included two stages. The first used a distinction method, identifying the program characteristics that distinguish the four programs that had evidence of favorable impacts from 9 programs that did not. The evaluation defined distinguishing characteristics as those that were present in at least three of the four

[5] The Third Annual Evaluation Report summarizes the qualitative findings from the implementation evaluation of the HCIA R2 awards. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

[6] This analysis excluded 5 awardees of the 19 for which it was possible to estimate program impacts. Four (AAMC, Altarum, Amerigroup, and FSCL) were excluded because the intervention was not expected to affect the core outcomes in the two- to three-year follow-up period of the study. One (Mesa) was excluded because of remaining concern that unobservable differences between treatment and comparison groups could lead to misestimation of true program effects in either direction.

programs with favorable impact estimates and in no more than four programs that did not have favorable estimated impacts.[7]

The second stage compared the median impact among awardees with a given distinguishing feature to the median impact among awardees that lacked the characteristic. This stage provides estimates of the magnitude of the difference in impacts between programs with versus without a given feature. The evaluation conducted assessments separately for each of the three core outcomes (total expenditures, hospitalizations, and ED visits).[8]

The analyses relied on 23 program features, organized into three broad categories: (1) 4 features related to the programs' intervention components, (2) 9 features related to program and awardee characteristics, and (3) 10 features related to implementation experience. The calculations were limited to features for which both the with- and the without-feature group contained at least three programs. **Appendix G** provides a full list of variables used to assess associations between program features and program impacts.

Of these 23 program features, the evaluation found 8 to be associated with favorable impacts in the distinction analysis (Table III.3), and for 7 of these the median analysis confirmed and quantified the association (Table III.4). That is, these 7 features were more common among the four programs classified as having evidence of favorable impacts than they were among the other programs, and the median impacts for programs with the feature were substantially more favorable than the median impacts for programs without the feature, for at least one outcome. The 7 features include 3 intervention components (behavioral health, telemedicine, and health IT) and 4 program design or awardee characteristics (having prior experience implementing similar programs, serving a socially fragmented population, focusing on individual patient care rather than transforming provider practice, and using nonclinical staff as frontline providers of the intervention). Programs using nonclinical staff as frontline providers of the intervention also had slightly more favorable median impacts on hospitalizations and ED visits than programs not using such staff. However, the difference between awardees that did or did not use nonclinical frontline staff was less pronounced than those found for other characteristics. Awardees that were hospital-based were more prevalent among the programs with evidence of favorable impacts, but the median estimated improvements in outcomes for hospital-based programs were actually smaller.

[7] UCSF was also excluded from the distinction comparisons because the sample size was too small to conclude the program was effective; although the impact estimates for all three core outcomes for this awardee were favorable, they were far from being statistically significant. This uncertainty made it potentially misleading to classify the program as either effective or ineffective. The calculation of median estimated impacts includes UCSF.

[8] The second stage analyses also compared the means of the impact estimates for awardees in each group, in addition to comparing the medians. The results were qualitatively similar for nearly all the features, but outlier values for some outcomes for one or two awardees sometimes distorted the means. Thus, the analysis relied on the comparison of medians.

Table III.3. Features distinguishing programs with versus without evidence of favorable impacts

Feature	Number of awardees with feature among:	
	Programs with evidence of favorable impacts (N = 4)	Programs without evidence of favorable impacts (N = 9)[a]
Program components		
Behavioral health	3	0
Telehealth	3	2
Health IT	3	3
Program and awardee characteristics		
Previous experience with similar programs	4	4
Socially fragile target population	4	4
Patient-focused intervention	3	4
Nonclinical frontline staff	3	4
Hospital-based program sponsor	3	4

[a] This analysis excluded six awardees without evidence of favorable effects: four (AAMC, Altarum, Amerigroup, and FSCL) because their interventions were unlikely to affect core outcomes during the observation period, one (Mesa) because of concerns about selection bias, and one (UCSF) because its sample size was too small to conclude the program was effective.

IT = information technology.

Table III.4. Median of percentage impacts on key outcomes for programs with and without a program feature

Feature (N with / N without)	Expenditures		Hospitalizations		ED visits	
	With feature	Without feature	With feature	Without feature	With feature	Without feature
Program components						
Behavioral health (3/11)	-12.5[a]	-1.0	-15.0[a]	1.0	-9.0[a]	-2.0
Telehealth (6/8)	-5.0[a]	0.0	-5.5[a]	1.5	-7.5[a]	-4.0
Health IT (6/8)	-8.0[a]	-1.0	-1.5	2.0	-7.5[a]	-4.5
Program and awardee characteristics						
Previous experience with similar programs (8/6)	-1.0	-4.5	-3.5	0.0	-8.5[a]	-1.0
Socially fragile target population (8/6)	-4.0	-1.5	-4.0	1.5	-7.5[a]	-1.5
Patient-focused intervention (8/6)	-6.5[a]	1.0	-6.5[a]	2.5	-6.5	-5.5
Nonclinical frontline staff (8/6)	0.0	-4.0	-2.5	-1.0	-7.0[a]	-4.5

Table III.4 (*continued*)

Feature (N with / N without)	Expenditures		Hospitalizations		ED visits	
	With feature	Without feature	With feature	Without feature	With feature	Without feature
Hospital-based program sponsor (7/7)	-2.5	-5.0	-1.0	-1.0	-6.0	-8.0

Note: This analysis excluded five awardees without evidence of favorable effects: four (AAMC, Altarum, Amerigroup, and FSCL) because their interventions were unlikely to affect core outcomes during the observation period, and one (Mesa) because of concerns about selection bias. The analysis excluded three additional awardees from the expenditure analysis (Montefiore, NACHRI, and NYC H+H) because they did not have data available on total Medicaid expenditures.

[a] Indicates favorable median impacts of a least -5% for programs with the feature, and at least 2.5 percentage points greater than median for those without feature.

ED = emergency department; IT = information technology.

The remainder of this section describes the associations found by the type of feature (intervention component, program or awardee characteristic, and implementation measure) and the likely mechanism by which the feature influenced outcomes. The discussion also explains the program features that appeared to affect outcomes and illustrates the findings with specific examples. It concludes by describing the implications of the findings.

1. Associations between intervention components and estimated impacts

Three intervention components were strongly associated with more favorable impact estimates. Those relationships were in general stronger than the associations between other program features and impacts. The overall median impacts (in percentage terms) across all 14 awardees included in the following analysis were -4.0, -1.0, and -6.5 percent for expenditures, hospitalizations, and ED visits, respectively.

Programs that relied on **integrating behavioral health services** with physical health services had substantially more favorable median estimated impacts on all three core outcomes. Behavioral health promotes well-being by recognizing that treating physical health problems often cannot be adequately addressed for some patients unless the patient receives adequate and coordinated treatment for mental illnesses such as depression or anxiety, or behavioral problems such as substance abuse and other addictions. UIC's CHECK program provides a good example of these programs. The program provided a range of mental health services to participants. These services evolved from educating participants during the first year to conducting regular mental health assessments, consulting with care coordination staff and participants' health care providers, and providing services and referrals starting in the second year. In the third year of the program, the awardee began offering expanded behavioral health services, such as one-on-one or online education in stress management and self-care.

The six awardees that relied on **telehealth** as a key intervention component also had more favorable median estimated effects on all three outcomes, especially for hospitalizations and total expenditures. Telehealth uses electronic information and telecommunication technologies to

support long-distance clinical health care, home monitoring through continuous sending of patient health data, patient and professional health-related education, meetings and presentations between practitioners, and online information and health data management. The distinguishing feature of the telehealth services provided by programs with evidence of favorable effects appears to be adoption of tools to monitor and respond to patients' needs as those needs arose. For example, as part of its eLTC program, Avera expanded its existing telehealth model to provide both transitional care coordination for skilled nursing facility residents and around-the-clock telehealth consults for skilled nursing facility residents and staff. Similarly, UCSF's care team navigators provided telephone-based support to link dementia patients and their caregivers with resources; provide supportive care and education; provide medication consultation; and support in planning future medical, financial, and legal decisions.[9]

Having **health IT** as a principal intervention component was also associated with more favorable estimated impacts, but not as strongly as the behavioral health or telehealth components. These programs' interventions relied heavily on computer hardware, software, or infrastructure to provide clinicians and other care providers ready access to critical patient information or treatment options and to share this information with other providers. For example, Montefiore's health IT component included a patient registry to collect and track participants' screening scores, between-visit follow-up communications, and participants' care plans and goals. Participants could also subscribe to an interactive voice response smartphone application that enabled them to complete follow-up monitoring measures via their phones, to receive appointment reminders and health education messages, and to communicate with patient educators. The increased amount and timeliness of information exchanges appear to have enabled programs to reduce patients' need for ED visits substantially.

2. Associations between program and awardee characteristics and estimated impacts

In addition to components of the intervention, other characteristics of the awardees and their programs can facilitate or impede their ability to favorably affect core outcomes. The evaluation identified five awardee characteristics in the distinction phase as being more prevalent among the programs with evidence of favorable effects than among those without such effects. For one of these characteristics—being a hospital-based awardee—median impact estimates were not more favorable than for awardees that were not hospital-based.

Awardees that had **previous experience** implementing an intervention like their HCIA R2 program had a substantially larger median estimated reduction in ED visits than those without prior experience. However, having previous experience did not appear to affect program impacts on hospitalizations or expenditures. It is likely that prior experience facilitated planning and early identification and amelioration of potential implementation barriers, such as establishing

[9] Telehealth differs from telemedicine in that it refers to a broader scope of remote health care services. Telemedicine refers specifically to remote clinical services, whereas telehealth includes these services and can include remote nonclinical services, such as provider distance-learning; meetings, supervision, and presentations between practitioners; online information and health data management; patient education; and remote admissions.

strategic partnerships and addressing staffing needs and requirements. Staff with greater experience with change might have been more comfortable adapting their workflows to accommodate innovations. Prior experience might also facilitate buy-in among providers and other factors associated with delivering quality service. For example, most of Montefiore's implementing practices had prior experience providing on-site integrated behavioral health. They also had implemented a measurement-based approach to care that enabled primary care physicians and behavioral health staff to work together in the primary care setting to provide behavioral health care services and referrals. In contrast, CCNC sought to implement a novel value-based approach in which pharmacists moved from only filling and dispensing medication to providing enhanced services to improve medication management for the most at-risk patients. The intervention required pharmacies to make major workflow changes and use a burdensome health IT system.

The awardees that served predominantly **socially fragile populations** had more favorable estimated impacts on all three core outcomes—especially ED visits—than the awardees that did not serve such populations. Socially fragile populations are those at higher risk for disease progression because of social circumstances or barriers. For example, homeless and indigent populations, people with language barriers or transportation issues, or people with adherence problems are considered socially fragile. One of the awardees with evidence of favorable impacts (NYC H+H) serves as the public safety net in the New York City's health care system. It provided ED care management services to high users of its emergency care services.

The programs whose interventions were primarily **patient-focused** also had substantially more favorable median estimated effects on total expenditures and hospitalizations than the provider- or facility-focused interventions. These programs might have been more effective on average because they addressed the barriers that individual patients faced to reducing their need for extensive and expensive care, rather than trying to change providers' behaviors.

Finally, the programs that relied on **nonclinical frontline staff**, such as community health workers and social workers, had more favorable median estimated effects on ED visits than the programs that had clinical staff in this role. These results were less robust than the previous results—median estimated reductions were only slightly more favorable for programs relying on nonclinical staff. The more favorable impact estimates might be due to the ability of social workers or community support providers to address the nonmedical problems that prevent patients from adhering to physicians' recommendations. The problems could include difficulties such as obtaining transportation to follow-up medical appointments or access to affordable medications. UIC represents an awardee using nonclinical staff on the front line of its intervention. Community health workers conducted initial assessments of participants to identify how connections to social service agencies might address their nonhealth needs and identified their physical and mental health needs. The community health workers then worked to connect participants and their families to relevant social service agencies and coordinate physical and mental health services, including the enhanced mental health services provided by the program.

Contrary to the stage one results, awardees that were **hospital-based** (versus community or other provider-based interventions) did not have more favorable median impacts, for any outcomes. The median estimated effects were similar for awardees with and without the feature, for all three outcomes.

3. Associations between implementation measures and estimated impacts

None of the ten measures of implementation effectiveness examined was meaningfully or consistently associated with more favorable median estimated impacts (not shown in table). This association could be due to the greater availability of management and staff attention to the intervention itself if they did not have to devote considerable time and resources to reaching enrollment targets. For the other measures examined, programs with the stronger rating on implementation had *worse* estimated impacts or no association at all.

This lack of association between stronger implementation and more favorable average impacts is consistent with the results from the distinction method, which showed no association between these features and whether a program had strong evidence of favorable effects. Measuring implementation is fraught with difficulties, including differences across programs in the standards used in assessing their own challenges and performance, changes over the course of the program in these measures, and the need to rely on respondents rather than the research team's direct observations. Furthermore, the degree to which an implementation factor affects program impacts varies widely across interventions. Thus, the evaluation cannot draw inferences about the associations between implementation effectiveness and program impacts.

4. Implications of associations between program features and impacts

All four programs with favorable estimated impacts, and none of the 10 programs without such evidence, shared three distinct features. They were:

1. Serving a socially fragile target population
2. Having pre-HCIA R2 experience with the intervention
3. Addressing participants' nonmedical needs, either by including a behavioral health component or relying on nonclinical staff to deliver intervention services

Furthermore, among the four programs with favorable impacts, all except NYCH+H had all of the intervention components identified as related to favorable effects (telehealth, health IT, and behavioral health). None of the programs without impacts had this combination of features. However, NYCH+H (a program with evidence of favorable effects) had none of these intervention components, indicating that these components are not essential for program success. NYCH+H was able to reduce hospitalizations and ED visits by relying on nonclinical staff such as social workers or community workers to address participants' social needs. Finally, programs that focused directly on addressing beneficiaries' needs had substantially higher median impacts than programs that tried to change providers' behavior.

These results do not imply that other factors were unimportant to achieving favorable impacts. However, other important factors were often present in programs that were not effective, as well as in programs that were effective. For example, having strong leadership and hiring highly competent and motivated staff are both likely to be important for a program's ability to reduce preventable hospitalizations or ED visits, but many programs that did not have evidence of favorable effects also had strong leadership and staff. Similarly, implementing an intervention well is clearly important for achieving the desired outcomes, but it is difficult to measure the effectiveness of implementation, and to determine which combination of implementation factors are critical for a given intervention's success.

These results also suggest that programs with many of these 7 features (3 of the 4 successful programs had at least 6 of them) might have a greater likelihood of more quickly achieving CMS's goals of reducing Medicare and Medicaid enrollees' needs for costly health care services. Socially fragile patients need help overcoming social barriers, such as lack of transportation and stable housing, to improve their health and reduce their need for hospital inpatient or outpatient services. Program staff such as community health workers and social workers are experienced in helping patients identify and access available social services and supports. These beneficiaries often have nonphysical problems such as depression or anxiety, or behavioral problems such as substance abuse and other addictions, which a behavioral health component can help to address. A program's experience with the intervention before testing it in a new area or group of patients also facilitates its ability to impact key outcomes quickly, shortening or eliminating the maturation process typically required for a new program to become effective. Creative uses of health IT and telehealth tools can overcome communication barriers among providers and between providers and patients, increase beneficiaries' timely access to specialists, and reduce beneficiaries' need for emergency room care. Although these inferences are drawn from a small set of programs, they align with theories of action and the literature as key contributors to successful interventions. The results from this study do not imply that only programs with these features can be successful in reducing health care expenditures and preventable use of expensive services, but such programs appear to increase the likelihood of early success.

IV. ASSESSMENT OF PROGRAM SUSTAINABILITY AND PROPOSED PAYMENT MODELS

Throughout their awards, awardees planned for how they could sustain their programs after HCIA R2 funding ended. For the second round of the HCIA initiative, CMMI added the expectation that awardees would develop a payment model that could generate ongoing funding for their programs and, if possible, test and potentially implement that model with payers by the end of the three-year cooperative agreements (August 2017). In doing so, CMS's goals were to (1) have awardees design models that could provide a sustainable source of funding for service delivery after the cooperative agreement ended and (2) be able to identify promising models that could inform the development and implementation of future models.

This chapter highlights findings in two areas: (1) the extent to which awardees sustained their programs after the award period, the sources of funding they used, and the factors affecting their ability to sustain their programs; and (2) awardees' status in implementing their proposed payment models, the features of the payment models most likely to be implemented, and the factors affecting implementation of the payment models. This qualitative assessment is based on awardee interviews conducted in late 2017 and mid- 2018, as well as quarterly awardee progress reports. Table H.1 in **Appendix H** summarizes the extent to which awardees sustained their programs, their proposed payment models, and the source of funding they used to sustain their programs, when applicable.

A. Findings from assessment of program sustainability

Overall, awardees made considerable efforts to continue at least parts of their programs after the award period, and many (29) were able to sustain at least parts of their programs. A few common factors helped awardees obtain the funding necessary to continue program components.

1. Program sustainment status

At the end of their awards, 11 awardees reported they were sustaining their entire programs (Table IV.1). Most of these continued programs focused on the Medicare population or both Medicare and Medicaid populations. These programs commonly addressed chronic conditions, although the range of services each program provided differed widely. For example, FSCL's program focused on increasing quality of life for people with life-limiting illnesses through community-based palliative care; Northwell provided care management services to patients with late-stage kidney disease; and CHIIC provided health coaching for patients with obesity and diabetes in rural, low-income communities. Three other sustained programs focused on acute and sub-acute conditions, such as providing long-term care in nursing facilities (Avera), telehealth consultations between physicians treating patients with neuro-emergent conditions in the ED (UNM), and respite care for homeless people after hospitalization (NHCHC).

Table IV.1. Program sustainability status by awardee

Continued program largely intact (N = 11)	Sustained parts of program (N = 18)	Ended program (N = 9)
Medicare		
AAMC, Avera, CHIIC, FSCL, Northwell, UKS, UNM	CCC, CCNC, Icahn, NMC, UCSF, UHCMC, UMich, Yale	ACCF, Hopkins, UNC, Ventura
Medicaid		
Montefiore, NHCHC, UIC	Altarum, Amerigroup, CHS, Clifford Beers, Columbia, DMC, NACHRI, Wash U, WI DHS,	BMC, SCH
Medicare and Medicaid		
FPHNY	NYC H+H	Mesa, UCSD, VCC

Source: Interviews and other ad hoc communication with awardees conducted in late 2017 and mid-2019; quarterly awardee-submitted reports

The fully sustained programs shared several attributes that were less prevalent in non-sustained programs. First, they were more likely to offer telehealth services. Second, they were more likely to have been characterized as effectively enrolling participants during the award period, meaning the awardee reached at least 90 percent of the enrollment goal it set at the beginning of the award period. Third, they were more likely to operate their program in rural communities, rather than urban or suburban areas.

Awardees commonly made minor modifications to their programs to sustain them. These changes were aimed at improving program efficiency and effectiveness without changing the core features of the program—for example, by aligning the program with other initiatives to reduce administrative burden or improving processes to more effectively reach program goals. Awardees also made modifications to help their programs fit with objectives of ongoing funding sources they needed to sustain the program. For example, during the award period, UKS's program focused on individuals with chronic conditions related to heart attacks and strokes, but the sustained version of the program encompasses individuals with a wider range of chronic conditions. By providing health coaching services to a wider range of patients, the program aligned with the goals of its Medicaid Shared Savings Program (MSSP) Accountable Care Organization (ACO).

The 11 awardees that continued their full programs relied on a range of funding sources to do so. Six of these awardees (Avera, CHIIC, Montefiore, NHCHC, UKS, and UNM) were able to implement at least part of their proposed payment model to fund their entire programs on an ongoing basis. In contrast, the other five awardees that sustained their programs had to find additional sources of funding to supplement funding from their payment models. The other sources of funding were typically provided internally by the implementing organization, but in some cases included external sources that could be time-limited, such as funds from participating providers, or grants from foundations, state or federal agencies. The most common types of implemented payment arrangements included FFS billing and PBPM payments; a few awardees

also had shared savings arrangements through the ACOs they already participated in. (See section B for more details on awardees' payment models.)

Eighteen awardees had funding to continue only parts of their programs (Table IV.1). For half of these awardees, partial sustainability meant making the program available to a subset of the original target population, either by narrowing the eligibility criteria, ceasing new enrollment, or limiting the number of participating providers. Other awardees decided to continue only certain program services. These awardees typically prioritized continuing their direct medical, dental, or behavioral health services, while ceasing to offer some of the support services, such as patient education, provider training and technical assistance, patient outreach and navigation assistance, and spiritual services. For instance, CHS closed one of its implementing sites, which meant ending its community-based patient navigation and outreach services. A few awardees with multiple participating sites (CCC, NACHRI, and NYC H+H) stopped managing their programs, and allowed the individual sites to make their own decisions about which components to continue.

Several awardees also expanded their programs to other sites or populations

- Avera, CHIIC, Montefiore, and UKS implemented their programs in additional provider locations
- AAMC and UHCMC expanded their programs to include different types of providers
- Altarum broadened the populations served by its program
- AAMC, Altarum, and CCNC replicated their programs in other communities

Nine awardees ended their programs due to lack of funding (Table IV.1). The terminated programs had targeted different populations, but they were more likely to focus on patients with chronic conditions compared to programs that were continued. Several of the awardees that ended their programs still maintained some of their infrastructure, such as the health IT resources that UNC had made available to participating sites. Also, awardees did not necessarily give up on their program concepts. For example, Ventura reported that its program inspired several new related initiatives, while Hopkins, VCC, and UCSD continued discussions with payers in hopes of implementing payment models to restart their programs.

2. Factors affecting sustainment

Three factors helped awardees obtain the necessary funding to sustain their programs in full or in part: (1) buy-in from leaders and staff, (2) delegating responsibility to participating sites, and (3) aligning the program with other initiatives. First, awardees that gained strong buy-in from leaders and frontline staff for their programs had an easier time building a case for sustaining the program. For example, AAMC leaders perceived that the eConsult function central to its program was more efficient than treating patients through traditional in-person visits or phone calls. As a result, they were willing to fund the program with internal resources until they could execute their proposed payment model. In addition, when practitioners and staff found value in

program components, they were more likely to incorporate and embed them into workflows in a manner that made them routine and enduring, and to encourage leaders to provide ongoing support. For example, because the implementing clinicians believed the FPHNY program improved quality of care, they integrated its core components of care coordination and telementoring into their regular practice and expressed concerns that the termination of these services would adversely affect patient care; the awardee decided to sustain the program using internal funds.

Second, delegating responsibility and flexibility to participating sites helped sustain programs. Several awardees that stopped overseeing their programs, while encouraging their participating sites to tailor the intervention to meet their own circumstances and needs, were able to sustain their programs in whole or part. These awardees typically assisted the sites in generating ongoing funding through a payment model or other source, but in some cases delegated that responsibility to local sites as well. For example, provider sites that participated in the NACHRI and NHCHC programs implemented their own payment models to fit the priorities of their state Medicaid programs as well as the characteristics of the program at those sites.

Factors affecting program sustainability

- Generating buy-in from leaders and frontline staff
- Delegating responsibility with flexibility to participating sites
- Aligning program aims with state and federal policies

Third, aligning programs with state or federal initiatives helped awardees access existing funds to sustain their programs, at least in the short term. Two of the programs based in New York (Montefiore and FPHNY) focused on similar areas as the state's Medicaid Delivery System Reform Incentive Payment (DSRIP) demonstration (ED care management and integration of primary and behavioral health care), and aligned some of the requirements so they could receive ongoing funding for them. CHS received funding from legislative appropriations from the Florida legislature to sustain parts of its school-based health services program. Also, UCSF received a 5-year grant from the National Institute on Aging to sustain part of its program to treat dementia.

B. Findings from assessment of proposed payment models

Most (36) awardees proposed a payment model at the beginning of their awards with the intention of using the funding generated by these models to sustain their programs after the award period. The awardees that proposed a payment model achieved different amounts of progress, and by the end of the award fell evenly into one of three categories: (1) those that implemented at least part of their payment models, (2) those that were in active discussions or negotiations with payers but had not yet entered contracts, and (3) those whose efforts to gain payer interest had stalled and appeared unlikely to proceed. Two awardees did not propose a payment model because they secured other funding early in the award period that would extend

past the award period to help sustain their programs. Amerigroup received a contract from the state to keep operating its program, and DMC used part of an existing revenue stream (although these funding sources proved insufficient to sustain their whole programs).

Twelve awardees succeeded in implementing at least part of their payment models. That is, they were able to implement at least one of the proposed payment arrangements with at least one type of plan or payer. Overall, these 12 awardees implemented 19 arrangements mostly with Medicare or Medicaid. However, one of these arrangements was with a commercial payer and another arrangement involved direct payments from providers (Table IV.2). The awardees that could not secure enough funding through their payment models before the award ended either continued only part of their program (six awardees) or found additional sources of funding (two awardees).

The most common payment arrangements awardees implemented were FFS billing codes used to obtain reimbursement from payers for services provided by the program (Table IV.2). Eight awardees used FFS billing codes to sustain their programs. Most of these codes were already established but had not been previously used by the awardee, while others were new. For example, CHIIC and UKS started using Medicare transitional care management (TCM) billing codes. Using FFS codes was easier than pursuing new and more complex arrangements, which required extensive negotiations with or changes from payers. In fact, several awardees who initially proposed new types of payment arrangements ultimately changed their payment approaches to FFS billing codes based on these challenges. For instance, UMich changed its payment model to use FFS codes, after having problems engaging surgeons with the performance incentives it originally proposed.

The remaining payment arrangements implemented by awardees focused on receiving payments for a fixed period of time (for example, per month) and/or a given set of services (for example, care coordination), including reimbursing services and activities that were previously not reimbursed by payers. Six awardees implemented a capitated PBPM payment. For Montefiore, NACHRI, UHCMC, and WI DHS, these were care coordination or care management fees. The other two awardees' PBPMs covered a broader set of services. Nursing facilities agreed to pay Avera a monthly fee for providing telehealth services (including training, care coordination, and urgent and specialty care consults) in its long-term care settings, while the PBPM fee for NHCHC covers a range of respite care services. Finally, Icahn successfully executed contracts with two commercial payers using a bundled payment model that covers inpatient and palliative care services that can be tailored to match payer preferences.

In addition, four awardees (CHIIC, Montefiore, NACHRI, and UKS) had executed shared savings arrangements, typically through a MSSP or Medicaid ACO of which they were a member. They planned to use these savings to support their programs over time.

Table IV.2. Characteristics of payment arrangements implemented by awardees

Source of payment and (number of arrangements)	Fee-for-service payment	Bundled or episode-based payment	Per beneficiary per month payment	Payment with shared savings
Medicare (9)	CHIIC, UHCMC, UKS, UNM		Montefiore, UHCMC	CHIIC, Montefiore, UKS
Medicaid (6)	NACHRI, NHCHC		NACHRI, NHCHC, WI DHS	NACHRI
Commercial (3)	Clifford Beers, UMich	Icahn		
Provider (1)			Avera	
Total (19)	**8**	**1**	**6**	**4**

Note: Because many awardees implemented more than one type of payment model and with multiple payers, the totals in these columns could be greater than the total number of awardees.

The payment models that lagged or stalled tended to be the more advanced, value-based payment arrangements. Many of the awardees (17 of 24) that had not implemented their payment models by the end of the award focused on alternative payment models without a FFS component. For example, CCNC continued to pursue a payment model that combined a monthly Medicare payment for care plans (the amounts of which would vary based on the patient's risk score and the pharmacy's performance on quality), with PBPM payments for care management and coordination services. Such arrangements involved more financial risk to the providers because their payments were adjusted for performance, but they also provided more flexibility to implement innovative aspects of programs that were not currently reimbursed. Because these arrangements were newer and more complex, they involved more discussions and negotiations with payers, additional data and analyses, and extensive process or policy changes. Several awardees reported ongoing negotiations with Medicaid MCOs for their payment models, which could mean the process took longer but remained promising. For example, FPHNY continued to pursue a one-time bundled payment from Medicaid and Medicare managed care plans to fund care coordination services, subject to shared savings and losses.

How the four programs with evidence of favorable impacts sustained their programs

- Avera used a PBPM payment model with payment from its participating nursing facilities
- Montefiore used a shared savings model with a Medicare ACO and a value-based PBPM payment model with other payers, plus DSRIP funding from New York State
- NYC H+H also relied on DSRIP funding from New York State
- UIC used internal funding while continuing to negotiate with its Medicaid MCOs

The biggest challenge awardees faced in implementing their payment models was lack of data for demonstrating the effectiveness of their programs and setting payment rates. Without this information, it was difficult to engage payers. Awardees

needed to calculate payment rates that covered the program's costs while also being acceptable to payers. For example, ACCF had difficulty accessing the claims data needed to determine the underlying costs of the program and therefore could not arrive at an appropriate payment amount. The awardee ultimately ceased operating its program due to lack of funding, as well as other challenges.

In addition, awardees struggled to engage payers without having evidence from the federal evaluation. Although many awardees tried to conduct their own analyses, they typically found that they needed more time, both to obtain sufficient claims and other data and for the program to lead to a discernible reduction in health care service use and costs. For instance, Altarum abandoned its payment model after it could not show that its preventive oral and dental health services program generated savings for Medicaid. Without being able to prove cost savings, the awardee could not justify paying providers incentives for meeting preventive service targets.

Several awardees reported being hopeful that evidence from the federal evaluation or internal studies based on more years of claims data would eventually reignite conversations with payers and lead to the implementation of their payment models in the future. In addition, a few awardees anticipated that the broader state and national movements toward value-based purchasing would ultimately support longer-term sustainability and potential replication of their programs, even if they could not reach agreements with payers before the award ended. For example, CCNC expected to not only incorporate its program into the state's upcoming Medicaid reform effort, but also that other Medicaid programs would work value-based purchasing arrangements into waivers and state plan amendments to help solve challenges similar to those the awardee was attempting to address with its program.

This page has been left blank for double-sided copying.

V. IMPLICATIONS OF THE FINDINGS

CMMI intentionally designed the second round of HCIA funding (like the first) to test a range of interventions. Its goal was to identify promising interventions that merit expanding, testing further, or incorporating into existing health care delivery systems. The HCIA R2 initiative added developing (and testing, if possible) innovative payment models that could help cover the cost of these and similar service delivery programs that have the potential to lower costs and improve care. The HCIA initiatives differed from most CMMI models because the awardees themselves developed the interventions. Therefore, the initiatives serve as a complementary source of innovative ideas to CMS's other more centrally designed and administered model tests. Given the range and innovative nature of the interventions that HCIA R2 tested, it was reasonable to expect that some programs would achieve promising impacts on health care costs, service use, and quality during the study period, whereas others would not. A key goal of this evaluation was to identify interventions that produced promising results based on the available data and describe future implications for promoting innovation in health care delivery and payment.

A. Summary of HCIA R2 evaluation findings

This evaluation identified 4 distinct interventions (among 19 for which it was possible to produce credible impact estimates) that had evidence of promising effects on spending and use of acute care services. Avera's eLTC program provided telehealth services to both short- and long-term residents in nursing facilities. It significantly reduced Medicare spending and ED service use. Montefiore's BHIP provided integrated behavioral health services in the primary care setting for adults and children with behavioral health disorders. The BHIP led to a significant reduction in ED service use and had a large estimated reduction in hospital admissions (expenditure data were not available for this awardee). UIC's CHECK program coordinated medical, nonmedical, and mental health services for children, specifically those with complex medical conditions. It led to a statistically significant reduction in Medicaid spending by reducing hospital inpatient and ED service use. NYC H+H's ED care management program provided care management in the ED and transitional care coordination for up to 90 days after discharge to help patients keep their follow-up physician appointments. It significantly reduced ED visits and hospital admissions (expenditure data for eligible Medicaid enrollees were not available for this enrollee).

The remaining programs funded under HCIA R2 were either not evaluable in a way that could produce credible impact estimates (19) or were evaluable but had estimates that were small, unfavorable, or too imprecise to support a conclusion of favorable effects (15).

B. Lessons learned from evaluation findings of HCIA R2 awardees

The results of the evaluation of the HCIA R2 awardees highlighted nine lessons with implications for implementing programs to promote innovation in health care delivery and payment models in the future. Even though there was considerable investment in the innovations, there is little evidence showing the potential for savings to offset the outlay of federal resources

under HCIA R2. Only 2 programs of the 19 evaluable programs showed a statistically significant reduction in health care expenditures; 2 other programs showed a reduction in service use, but lack of Medicaid data prevented an examination of their impact on spending. Second, although the purpose of the cooperative agreement was to support and facilitate investigator-defined service delivery and payment reform models, the cooperative agreement approach does not lend itself to evaluating the impact of implementation of changes in payment policy.

Third, even though many awardees saw value in terms of improved patient care in continuing their programs, lack of success negotiating contracts with payers highlights the challenges of sustaining innovations in delivering health care. Funding innovative programs is particularly difficult within an FFS payment environment that does not reimburse providers for many of the enhanced services offered under the awards. Furthermore, within a model in which those paying for the intervention would reap some or all of any savings achieved, awardees had difficulty convincing payers such as MCOs or ACOs to pay for their services without stronger evidence of expected savings. Although the awardees with favorable impacts sustained their programs, only one (Avera) could

Lessons learned from the HCIA R2 evaluation

- Even though there was considerable investment in the innovations, there is little evidence showing the potential for savings to offset the outlay of federal resources under HCIA R2.
- Although the purpose of the cooperative agreement was to support and facilitate investigator-defined service delivery and payment models, the cooperative agreement approach does not lend itself to evaluating the impact of implementation of changes in payment policy.
- Despite perceived value among clinicians, convincing payers and plans to cover the cost of innovative programs is difficult within an FFS environment and without clearer evidence that savings to the payer are likely to exceed the cost.
- It is difficult to move the needle on health care delivery reform in three years, particularly among patients with chronic conditions.
- It is easier to reduce ED visits than it is to reduce hospitalizations.
- Intervention effects tend to be concentrated among subgroups of higher-risk beneficiaries most likely to benefit from enhanced services.
- Even when programs achieve favorable results, it is difficult to produce sizeable savings, making it difficult to cover or exceed the cost of the intervention within a limited time period.
- Very different programs can achieve favorable results, suggesting there are many opportunities to deliver better care.
- Interventions that targeted a socially fragile population, had previous experience, and either had a behavioral health component or used nonclinical staff to work with participants had favorable impacts, whereas those without these 3 features did not. Programs that relied on telehealth and/or health IT also had substantially larger median impacts than those without these features. Having a patient-focused or a hospital-based intervention was more weakly linked to more favorable impacts.

rely completely on one of its two proposed payment models, further highlighting the complexity of negotiating new models. Most of the awardees that lacked evidence of savings sustained their programs in whole or part at the end of the award with the help of internal resources, funds from existing FFS billing codes, or external grants. Many of these awardees could continue their programs because they had buy-in from organizational leaders and frontline staff, had delegated responsibility and flexibility to the sites participating in the program, and had aligned the aims of their program with state and federal policies. However, whether awardees can continue their programs over the longer term is uncertain.

Fourth, it is very difficult to make considerable advances in reforming the delivery of health care in three years. This is especially true when part of that time focuses on starting the program. Programs attempting to improve care and encourage behavioral change among patients with chronic conditions are likely to take more time to have favorable effects on health care costs and service use. The evaluation's finding of an association between previous experience implementing a similar program with a similar patient population and achieving favorable impacts supports this conclusion.

Fifth, the impact findings suggest that it is easier to reduce ED visits than it is to reduce hospitalizations. However, reductions in ED visits are unlikely to achieve the same level of cost savings as reductions in hospitalizations. In addition, interventions that reduce ED visits are likely to be more effective for Medicaid beneficiaries who often have competing social and economic priorities and are less likely to have a regular source of care than Medicare beneficiaries.

Sixth, program effects tend to be concentrated among subgroups of high-risk Medicare and Medicaid participants. The benefits of Avera's program were concentrated among long-term care nursing facility residents, the benefits of NYC H+H's program were limited to early enrollees who tended to be sicker and to have greater service needs than those who enrolled later, and the benefits of UIC's program were limited to higher-risk children who had one or more ED visits or one or more inpatient hospitalizations in the year before random assignment. This finding is consistent with the existing literature on the impacts of care coordination interventions and demonstrates the importance of targeting delivery system innovations at subpopulations at greatest risk for preventable service use (Brown et al. 2012).

Seventh, even when a program achieves positive results, it is difficult to produce large savings from health care delivery innovations, making them unlikely to cover or exceed the cost of the intervention, especially within a limited time period. Only 1 program had sizeable and statistically significant reductions in spending (21 percent for UIC), and this estimate was inflated by outliers in the control group. Avera reduced costs, but by only 4 percent. Of the other 17 programs with impact evaluations, program impact on total expenditures could be estimated for 11 programs. Of those, 7 had estimated cost reductions, but the estimates were generally small and too imprecise to be statistically significant.

Eighth, very different types of programs can achieve favorable results, suggesting there are many opportunities to deliver better care. The effective programs differed on a number of key dimensions: who they served, the type of intervention they implemented, and the setting within which the awardee implemented the program. Of the four awardees with favorable impacts, one served Medicare residents in nursing facilities, one served Medicare and Medicaid patients discharged from EDs, one served Medicaid children with complex medical conditions in the community, and one served Medicaid adults with behavioral health disorders in primary care practices. The findings from this study suggest there are several types of people for whom programs can find ways to improve care, and various ways to accomplish it.

Ninth, despite the diversity of the programs with evidence favorable effects, they shared a few characteristics that were less evident among programs without evidence of favorable effects. All four programs with favorable effects (compared with just under half of the programs without favorable effects) had prior experience implementing similar programs and targeted a socially fragile population. A few of the programs without evidence of favorable effects also shared these characteristics. However, only the 4 programs with evidence of favorable effects also met the nonmedical needs of their socially fragile participants by having either a behavioral health component or by using nonclinical staff to work directly with the participants. Behavioral health interventions can help these patients address nonphysical problems they often face, such as depression or anxiety or substance abuse, while nonclinical staff such as community health workers and social workers are experienced in helping patients identify and access available social services and supports. In addition, most of the programs with evidence of favorable effects relied on telehealth and health information technology in delivering their interventions and focused on meeting patients' needs rather than on changing providers' behavior. These results do not imply that only programs with these features can be successful in reducing health care expenditures and service use, but such features appear to increase the likelihood of early success.

C. Final remarks

The two rounds of HCIA funding represented a novel approach to testing a diverse set of health care interventions that aimed to improve quality and efficiency of care. While the pursuit (and evaluation) of locally defined health care innovations is inherently challenging, CMS succeeded in identifying four very different programs with favorable results on one or more of the core outcomes of this study. Despite their differences, these promising programs shared a unique set of charactertistics, including targeting a socially fragile population and focusing on their nonmedical needs by either including a behavioral health component or relying on nonclinical frontline staff to deliver services. Many awardees also found value in their programs; three-quarters of the awardees sustained their programs, either in whole or in part, after the end of the award period. Although convincing payers to cover innovative health care interventions can be difficult without clearer evidence of savings, the continued movement away from FFS and toward paying for value and population health could help to facilitate additional support for health care delivery innovation in the future.

REFERENCES

Brown, Randall, Deborah Peikes, Greg Peterson, Jennifer Schore, and Carol Razafindrakoto. "Six Features of Medicare Coordinated Care Demonstration Programs that Cut Hospital Admissions of High-Risk Patients." Health Affairs, vol. 31, no. 6, June 2012, pp. 1156-1166.

Gelman, A. "Analysis of Cariance—Why It Is More Important than Ever." *The Annals of* Statistics*,* vol. 33, no. 1, 2005, pp. 1–53.

Gelman, A., J.B. Carlin, H.S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. New York: Chapman and Hall, 2004.

Hirano, K., and G. Imbens. "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services and Outcome Research Methodology,* vol. 2, 2001, pp. 259–278.

Pimentel, S.D., L.V. Forrow, J. Gellar, and J. Li. "Optimal Matching Approaches in Health Policy Evaluations Under Rolling Enrolment." *Journal of the Royal Statistical Society: Series A (Statistics in Society),* online ahead of print, 2019. Available at https://www.mathematica.org/our-publications-and-findings/publications/optimal-matching-approaches-in-health-policy-evaluations-under-rolling-enrolment.

Rosenbaum, P.R. "Optimal Matching for Observational Studies." *Journal of the American Statistical* Association*,* vol. 84, no. 408, 1989, pp. 1024–1032. doi:10.1080/01621459.1989.10478868.

Rubin, D.B. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology,* vol. 2, nos. 3-4, 2001, pp. 169–188.

Society of Actuaries. "2000–2011 Long-Term Care Aggregated Databases." Schaumburg, IL: Society of Actuaries, 2020. Available at https://www.soa.org/resources/experience-studies/2015/research-ltc-study-2000-11-aggregrated/. Accessed March 11, 2020.

Statistical Modeling, Causal Inference, and Social Science. "What Is the 'True Prior Distribution?' A Hard-Nosed Answer." Available at https://statmodeling.stat.columbia.edu/2016/04/23/what-is-the-true-prior-distribution-a-hard-nosed-answer/.

This page has been left blank for double-sided copying.

Appendix A:

Strategy for estimating program impacts

This page has been left blank for double-sided copying.

A. Overview

This appendix describes the core outcomes examined in the evaluation and the approaches used to estimate the counterfactual—that is, the outcomes expected for enrollees had the program not existed. The analysis used two different approaches to estimate this counterfactual. For most awardees, it used a difference-in-differences design that estimated impacts as the differences in outcomes between treatment and matched comparison beneficiaries after enrollment minus the estimated differences between the two groups in the pre-enrollment period. The other approach—a cross-sectional design—compared outcomes in the follow-up period, while controlling for baseline characteristics and values of outcome measures (which might have differed at baseline). The analysis used this second approach for awardees that were part of a randomized controlled trial (RCT) and for those for which the nature of the intervention did not support a difference-in-differences model because the baseline outcomes occurred before a major shift in a participant's health trajectory, meaning that the change over time in a beneficiary's outcomes was not a relevant indicator of the program's impact.

B. Estimation of program impacts

The impact evaluation assessed the impact of the programs on several key outcomes of interest that can be broadly grouped into three categories: (1) total expenditures (Medicare fee-for-service [FFS] and Medicaid FFS, when available), (2) service utilization, and (3) likelihood of acute care service utilization. The evaluation measured a common core set of outcome measures across awardees when applicable to the program (Table A.1). The evaluation also estimated effects on awardee-specific outcome measures that apply only to a given program. The data were constructed for each 3-month period after enrollment; continuous variables were then summed over adjacent quarters to create outcomes covering 6- or 12-month intervals used in analysis.

The analysis used ordinary least squares to estimate the models in Stata and estimated outcomes in 6-month periods from the beneficiary's enrollment date. However, the results presented in the body of this report are typically only for 12-month intervals, to reduce the variance in the outcome measures. The purpose for having results for 6-month intervals was to assess whether some minimum exposure to the intervention was necessary before program effects began to emerge. Such cases could mask program effects by estimating impacts over the first year or full period of program participation. The rest of this section describes the different approaches used to estimate the impact of the awardees' programs.

Table A.1. Core outcome measures used in the evaluation

Variable	Description	Definition of the variable
Expenditures ($ PBPM)		
Total FFS expenditures	Total FFS expenditures calculated from all FFS claims for each participant with at least one eligible day during that quarter	The sum of total FFS expenditures during eligible days divided by the eligibility fraction for that quarter and divided by 3
Health care service use rates (per 1,000 beneficiaries)		
Acute hospital admissions	Number of acute care hospital admissions calculated from claims data for each participant with at least one eligible day during that quarter	The sum of all hospital admissions during the quarter divided by the eligibility fraction for that quarter, multiplied by 4, and then multiplied by 1,000
Outpatient ED visits and observation stays	Number of outpatient ED visits and observation stays calculated from claims data for each participant with at least one eligible day during that quarter; this excludes observation stays and ED visits that lead to hospitalization	The sum of all outpatient ED visits and observation stays during the quarter divided by the eligibility fraction for that quarter, multiplied by 4, and then multiplied by 1,000
Measures of any health care service use		
Percentage of participants with a hospital admission	Percentage of participants with an acute care hospital admission for participants with at least one eligible day during that quarter	This variable takes the value 1 for every participant who had at least one hospital admission during that quarter and 0 if not
Percentage of participants with an outpatient ED visit or observation stay	Percentage of participants with an outpatient ED visit or observation stay, which excludes any ED visit that leads to a hospitalization, for participants with at least one eligible day during that quarter (or full observation period)	This variable takes the value 1 for every participant who had at least one outpatient ED visit or observation stay during that quarter (or full observation period) and 0 if not
Percentage of discharges with a 30-day unplanned readmission among all eligible index discharges[a]	Percentage of discharges with a 30-day unplanned readmission; indicator variable (0/1) defined at the discharge level	At the discharge level, this variable takes the value 1 if there was at least one or more subsequent readmissions within 30 days of the discharge and 0 if not
Percentage of participants with a 30-day readmission among all sample members	Percentage of all participants with a 30-day readmission with at least one eligible day during that quarter	At the participant level, this variable takes the value 1 for every participant who had at least one or more readmissions within 30 days after discharge during that quarter and 0 if not

Note: All measures, except discharge-level 30-day readmissions, are weighted by a matching weight and an eligibility weight, the latter of which reflects part-year Medicare FFS and Medicaid eligibility. Measures of expenditures are per beneficiary per month. The expenditure measure is not price standardized. Measures of service utilization are annualized and are per 1,000 beneficiaries. Measures of any health care utilization reflect the likelihood of use.

[a] This measure definition is based on the Yale readmission measure developed by the Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE 2018) that is used in the Hospital Readmission Reduction Program under Section 3025 of the Affordable Care Act. Additional information about the Yale readmission measure is available at QualityNet, "Measure Methodology Reports: Readmissions Measures," at https://www.qualitynet.org/dcs/ContentServer?cid=1219069855841&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

1. Difference-in-differences models

The analysis used a difference-in-differences model to evaluate most programs. The difference-in-differences model estimates the changes in outcomes before and after enrollment for beneficiaries in the treatment group and for the matched comparison group over the same period. It estimated the impact of the program as the difference between the average change over time for treatment beneficiaries and the average change over time for the matched comparison beneficiaries. Impact estimates based on the difference-in-differences framework assume parallel trends for the treatment and comparison groups at baseline. That is, the difference-in-differences estimates are likely to be unbiased as long as there were no significant differences in outcome trends between the treatment and comparison groups at baseline, or reason to suspect that trends would differ for the two groups had the intervention not occurred.

a. Difference-in-differences with a longitudinal panel of patients

For most awardees, the evaluation used a longitudinal design, which used the same individuals in the pre-intervention and post-intervention periods. For each outcome, a single regression model included all pre- and post-enrollment observations that were available for each individual in the sample to estimate impacts jointly for all six-month intervals. Equation (1) specifies the regression model used to estimate the impact of the program for continuous and count outcomes such as total expenditures or the number of hospitalizations:

$$(1) \quad y_{it} = \alpha + b_i + \gamma_t * p_t + \pi_t * HCC_i * p_t + \theta_t * treatment_i * p_t + \mu_t * mature_i * p_t + \delta_t * treatment_i * mature_i * p_t + \varepsilon_{it}$$

Where y_{it} represents a claims-based outcome variable for beneficiary *i* in time period *t;* α is a constant term; b_i is a beneficiary-level fixed effect for beneficiary *i*, which controls for all time-invariant beneficiary characteristics; p_t (for "post") is an intervention-period indicator that takes the value of 1 during a specific intervention period, for instance, the first six-month period after enrollment, and 0 otherwise; HCC_i is the HCC score for beneficiary *i* measured at baseline; and $treatment_i$ is a binary indicator of intervention status; the indicator takes the value of 1 if beneficiary *i* is in the intervention group, and is otherwise 0. The main effect of this indicator is not identified in this equation since it is collinear with the beneficiary fixed effects. $mature_i$ is a binary indicator of program maturity at the time of enrollment for beneficiary *i*. The indicator takes the value of 1 if beneficiary *i* enrolled in the program after the program was considered to have reached a mature stage of implementation, and is otherwise 0. The main effect of this indicator is not identified in this equation since it is collinear with the beneficiary fixed effects.

ε_{it} is the idiosyncratic error term. It represents unexplained variability in the outcome variable for beneficiary *i* during period *t*.

The Greek letters are parameters to be estimated. For example, the intervention period-specific coefficients (γ_t) capture changes experienced by the comparison group between follow-up interval *t* and the baseline. The π_t coefficients capture the differential association between the baseline hierarchical condition category (HCC) score and the outcome in each follow-up period (because the HCC score is a period-specific measure of risk that predicts Medicare expenditures, the model allowed the relationship between the fixed baseline HCC score and the outcome to vary in early versus later follow-up periods). The θ_t coefficients are the interval-specific difference-in-differences impact estimates for beneficiaries who enrolled before the program matured. The impact estimates for beneficiaries in the mature cohort are obtained as the interval-specific sums of the θ_t and the δ_t coefficients.

The analysis used the same approach for estimating impacts over follow-up intervals differing in length—for example, in estimating yearly or cumulative impacts versus semiannual impacts. That is, even if outcomes were measured over different lengths of follow-up intervals (*t*), such as 12 or 24 months versus 6 months, the impact estimates were obtained using the same approach as in Equation (1), after aggregating the data to the time period for which impacts were estimated.

In the case of awardees that did not consider program maturity a critical element affecting the program's impact on beneficiaries' outcomes, the analysis estimated a simpler version of Equation (1) without the maturity interaction. In that specification, the θ_t coefficients are the interval-specific difference-in-differences impact estimates for all beneficiaries.

Estimating impacts on binary outcomes with longitudinal data on beneficiaries used a different approach. Because a difference-in-differences model relies on taking the first difference (outcome in post-period minus outcome in pre-period) for each beneficiary, with a binary outcome the first difference would be zero for beneficiaries who did not experience the outcome in either period and those who did experience the outcome in both periods. Therefore, the analysis of binary outcomes did not use the difference-in-differences approach and instead used an intervention period-only model. This model controlled for the baseline outcome and its interaction with treatment status to allow the treatment effect to vary with the value of the baseline outcome. Also, instead of beneficiary fixed effects, the analysis controlled for beneficiaries' characteristics at baseline. The baseline characteristics were typically those used to estimate the propensity score used to selected a matched comparison group; for Medicare beneficiaries, that score included at minimum demographics, original reason for Medicare entitlement, dual eligibility, HCC score, total expenditures, the number of ED visits, and the number of hospitalizations. For Medicaid awardees, the list typically included demographics and service utilization but tended to vary more based on data availability and the target population of the awardee.

The model included site or facility indicators if treatment and comparison cases were in multiple sites or facilities. Similar to Equation (1), models for binary outcomes included the interactions

of treatment and maturity status indicators with intervention interval indicators (starting from the second intervention interval), and also triple interaction terms for treatment, maturity, and each intervention interval. However, unlike in Equation (1), the binary outcome models included the main effects of treatment status and program maturity status in estimating effects on binary outcomes.

Outcomes such as 30-day readmissions observed at the discharge level instead of at the beneficiary level used the difference-in-differences approach. However, the model did not use beneficiary fixed effects, because the analysis did not necessarily have repeated observations on the same beneficiaries over time. Instead, the analysis controlled for beneficiaries' characteristics at baseline and discharge-level risk factors drawn from individual conditions included in the HCC algorithm (Table A.2), including the main effects of treatment status, program maturity status, indicators for follow-up period intervals, and their interactions.

Table A.2. Risk factors included as control variables in discharge-level readmission models

- Severe Infection
- Septicemia and sepsis
- Infectious diseases and pneumonias
- Metastatic cancer and acute leukemia
- Severe cancer
- Other cancers
- Diabetes mellitus
- Protein-calorie malnutrition
- Significant endocrine and metabolic disorders
- End-stage liver disease
- Rheumatoid arthritis and inflammatory connective tissue disease
- Severe hematological disorders
- Coagulation defects and other hematological disorders
- Iron deficiency or other unspecified anemias
- Drug or alcohol psychosis or dependence
- Psychiatric comorbidity
- Hemiplegia, paraplegia, paralysis
- Seizure disorders and convulsions
- Respirator dependence/tracheostomy
- Cardiorespiratory failure and shock
- Coronary atherosclerosis or angina
- Specified arrhythmias and other heart rhythm disorders
- Coronary obstructive pulmonary disease (COPD)
- Fibrosis of lung or other chronic lung disorders
- Transplants
- Dialysis
- Renal failure
- Decubitus ulcer or chronic skin ulcer
- Hip fracture or dislocation

Note: Conditions are derived from the conditions used to estimate the HCC risk score.
HCC = hierarchical condition category.

In all models, standard errors adjusted for clustering at the unique beneficiary level to allow for serial correlation of the outcomes of individual beneficiaries over time in our longitudinal data set. The models were also weighted by an analytic weight that is the product of the matching weights described in Appendix B and an eligibility weight. The matching weight equalized the contributions of each matched set of comparisons per treatment beneficiary and the eligibility weights accounted for the number of months the beneficiary was alive and enrolled in Medicare FFS or Medicaid from the start to the end of the period.

b. Difference-in-differences with repeated cross-sections of patients

For a small number of awardees—those for which the intervention was at the facility level (for example, a change in a practice's use of electronic health records) rather than the patient level—the impact analyses used repeated cross-sections of beneficiaries in baseline and the intervention period. In these cases, the difference-in-differences approach relied on controlling for beneficiaries' characteristics during the year before the period over which the outcomes were measured, instead of using beneficiary fixed effects, for all outcomes. Thus, beneficiaries who met the eligibility criteria during the year before the intervention formed the baseline period observations, and beneficiaries who met the eligibility criteria during the intervention period comprised the follow-up period observations. For both baseline and intervention period cases, the analysis controlled for the sample member's characteristics measured over the year before the period over which outcomes were measured. It also included indicators for each follow-up interval, interactions between the HCC score and follow-up intervals, an indicator for treatment status, and its interactions with follow-up intervals that capture the difference-in-differences impact estimate.

The analysis estimated heteroskedasticity-robust standard errors in some regressions to account for the likelihood that the variance of the error term was not the same across all observations. The analytic weight was constructed in the same way as in the longitudinal difference-and-differences models described earlier.

2. RCT and cross-sectional models

For RCTs the awardee randomly assigned each study participant into either a treatment group or a control group. The analysis could obtain unbiased estimate of program impacts by computing simple differences in the mean values of outcomes between the treatment and control groups. However, it estimated more precise impacts by including covariates and estimating regression models of the following form:

$$(2) \quad y_{it} = \alpha + \theta_t * treatment_i + \beta'.X_i + \tau' * C_i + \varepsilon_{it}$$

where y_{it} represents a claims-based outcome variable for beneficiary *i* in time period *t;* α is a constant term; $treatment_i$ is a binary indicator of intervention status; the indicator takes the value of 1 if beneficiary *i* is in the intervention group, and is otherwise 0; X_i are beneficiary characteristics such as gender, age, hierarchical condition category (HCC) scores, and other pre-enrollment characteristics, including baseline values of outcome measures; C_i are other characteristics that may affect outcomes such as community features (for example, number of primary care physicians in geographic area) or hospital characteristics (for example, occupancy rate). For awardees with multiple facilities or treatment sites, site indicators in equation (2) accounted for potential differences in beneficiary outcomes by site. ε_{it} is the idiosyncratic error

term. It represents unexplained variability in the outcome variable for beneficiary *i* during period *t*.

Equation (2) was estimated separately for each period, so estimates of all parameters (the Greek letters in the equation) were obtained for all periods $t = \{1,2,\ldots,P\}$. The key parameter of interest is θ_t, which measures the impact of the program in participants' *t*-th period after enrolling.

The analysis adopted a similar post period-only estimation approach for a few awardees without random assignment for which a difference-in-differences model did not seem appropriate, such as those awardees enrolling patients at the time of a precipitating event. For example, if a program enrolled only patients at the time of a hospital discharge or skilled nursing facility admission, enrollment will present an idiosyncratic spike in utilization that might violate the parallel trend assumption required for a difference-in-differences model. The mechanics of the impact estimation for an RCT is identical to that of a post-period-only cross-sectional design. The key difference is that in an RCT beneficiaries are randomized to treatment and comparison groups and in the cross-sectional design a set of matched beneficiaries not receiving the intervention serve as the comparison group. Assuming that external trends affect the treatment and comparison groups similarly, a comparison group well matched on observed characteristics and assumed to also be well matched on unobserved characteristics should produce unbiased estimates of program effects under a post-period cross-sectional design.

Similar to the difference-in-differences models, comparison group observations in non-RCT cross-sectional models were weighted by the product of the matching and the eligibility weights. For awardees that used RCT designs, the estimations included only the eligibility weights because a matched comparison group was not constructed.

C. Sensitivity analyses

Several sensitivity analyses that modified the main specification were used to verify the robustness of the impact estimates.

1. Extending the baseline period

The baseline period was extended to include the full two years before the enrollment date—as opposed to only one year as in the main impact analysis. Program impacts were estimated including the same covariates as in the main analysis. The preferred specification was setting the 12 months before the program as the baseline period. This approach measured impact estimates relative to a reference period immediately preceding the enrollment date, which could better capture the health status of beneficiaries at the start of the program. However, the sensitivity analysis provides insights into the robustness of the impact estimates to the length of the baseline period. If trends in outcomes for the treatment and comparison groups are not parallel during the baseline period, the impact estimates are likely to change substantially as the baseline period

extends back an additional year. In practice, increasing the length of the baseline period had little effect on the impact estimates in nearly all cases.

2. Top-coding outcomes

The analysis examined sensitivity of the results to outliers by top-coding outcome variables for both the treatment and comparison groups at the 98th percentile of the outcome distribution in the treatment group. That is, all values above the 98th percentile were replaced with the value of the outcome variable at the 98th percentile and then the models were estimated using the top-coded variables.

Appendix B:

Propensity score matching methodology

This page has been left blank for double-sided copying.

A. Overview

The evaluation selected nearly all of the comparison group beneficiaries who served as counterfactuals for the impact analysis using propensity score matching (PSM). Using PSM enabled the study team to construct a comparison group that was similar to the treatment group on key observable characteristics that affected treatment status and outcomes. The method uses logistic regression to estimate the probability that each treatment and potential comparison group beneficiary is actually in the treatment group. Each treatment group beneficiary is then matched to one or more beneficiary in the potential comparison pool with a similar predicted probability (propensity score) of being in the treatment group. The covariates used in the logistic model are referred to as matching variables. When the distribution of each matching variable is similar in the treatment and comparison groups, the groups exhibit good covariate balance, and the match is typically considered acceptable. The evaluation used four measures to assess the quality of the match:

1. **Standardized differences in means.** The evaluation defined the standardized difference as the treatment–comparison difference between the mean values of a covariate, expressed in standard-deviation units. Smaller standardized differences indicate more closely matched groups. A commonly invoked benchmark (Rubin 2001) suggests that groups are well matched if standardized differences for all covariates are less than 0.25. However, the evaluation strove for differences no larger than 0.10.

2. **Percentage difference in means.** Even when the standardized difference is less than 0.10, the percentage difference in means can be quite large, especially for variables with a high variance, such as baseline expenditures and number of hospitalizations. For example, a standardized difference of .10 for a variable with a coefficient of variation of 2.0 (a typical value for baseline expenditures and hospitalizations in the study's samples), the absolute difference in means would be 20 percent of the mean. A difference this large in baseline means would cause concerns about the assumption that the study drew treatment and comparison groups from the same population and would have similar outcome trajectories in the absence of the intervention. Thus, the study team also examined the absolute difference in means and used calipers and other matching methods to keep the percentage difference in means on key characteristics, such as baseline values of outcome variables, below 10 percent whenever possible.

3. **Equivalence tests.** An equivalence test expresses the null hypothesis as stating that the absolute value of the difference between two means is greater than a specified amount. The tests the study team conducted specified a difference in covariate means of at least 0.25 standard deviations. Rejecting the null hypothesis, implying that the difference in means is less than 0.25 standard deviations, suggests an adequate match.

4. **T-test for difference in means.** The study team also conducted a standard t-test for differences in the mean value of each covariate. Unlike the two previously described measures, the t-test is not typically recommended as a test for the quality of a match because samples that are large enough will often lead to rejection of the null hypothesis. In many

cases, the differences are statistically significant but operationally insignificant; the equivalence test described earlier assesses the latter, more relevant, criterion. Even more concerning, when samples are small, the test can fail to reject the hypothesis of equal means for the two groups even when the differences are large because the power of the test is low. The study team included this test because it could signal issues with the match that deserve further investigation. Note that the t-test and the equivalence test can both reject the null hypothesis (that is, the means are not equal for the two groups, but the difference does not exceed 0.25 standard deviations), especially when samples are large.

B. Detailed methodology

Although the study team used PSM to select comparison cases that best match treatment cases on propensity scores, techniques differ in how to assign treatment and comparison beneficiaries to one another based on their propensity scores. This section elaborates on the techniques used for Round 2 of the Health Care Innovation Awards (HCIA R2).

1. Default strategy

To achieve methodological consistency across the large set of awardees eligible for matching, the study team standardized the matching process as much as possible. Variable-ratio optimal matching with replacement at the beneficiary level was the default matching approach for all awardees.

Variable-ratio matching means that an analysis can match different treatment beneficiaries to different numbers of comparison beneficiaries, in contrast to fixed-ratio matching, in which the analysis matches each treatment beneficiary to the same number of comparison beneficiaries. With variable-ratio matching, treatment A could be matched to comparisons 1, 2, and 3, and treatment B is matched to comparison 4 only. Fixed-ratio matching can reduce the quality of the comparison group by forcing the sample to include poor matches; in the example above, fixed-ratio matching with a 1T:3C (one treatment to three control beneficiaries) ratio would require the study team to find additional comparison beneficiaries to match to treatment B. If those comparison beneficiaries are not particularly good matches, including them in the sample will lower the balance of the two groups—that is, the degree to which the comparison and treatment groups have similar means on included covariates. Variable-ratio matching remedies this pitfall by assigning more comparison beneficiaries to treatment beneficiaries with many strong candidate matches, without forcing additional poor matches. For HCIA R2, the study team allowed each treatment beneficiary to match to up to five comparison beneficiaries.

Optimal matching (Rosenbaum 1989) is an algorithm for assigning potential comparison beneficiaries to treatment beneficiaries. It is optimal in the sense that it selects matches to minimize the sum of the differences in propensity scores between treatment beneficiaries and their matched comparison beneficiaries across the entire sample. This process produces the best-matched overall comparison group, as opposed to nearest-neighbor or so-called greedy matching,

which focuses on finding the best possible match for each individual treatment beneficiary without considering the consequences for the sample as a whole.

Matching with replacement means that the study team allowed selecting the same comparison beneficiary as a match for more than one treatment beneficiary. The team did so to improve the quality of the comparison group when some members of the treatment group might be difficult to match. Suppose the analysis is matching on a single covariate, whether the beneficiary had high baseline Medicare expenditures, and there are three treatment beneficiaries with high baseline expenditures compared to only one potential comparison beneficiary in this category. Without replacement, the analysis will either exclude two treatment beneficiaries from the analytic sample or match two treatment beneficiaries with high baseline spending to comparison beneficiaries with low baseline spending. Matching with replacement allows the analysis to keep all the treatment beneficiaries in the sample without sacrificing the quality of the comparison group. Still, the analysis would prefer to limit the number of treatment beneficiaries to whom a comparison beneficiary can be matched, to minimize the effect of the matching weights on the variance of the impact estimates and for face validity. To that end, the study team allowed comparison beneficiaries to match to no more than five treatment beneficiaries.

The matching process produces matched sets that contain some number of treatment beneficiaries and some number of comparison beneficiaries. In the previous examples, a first matched set contains treatment A and comparisons 1–3, and a second contains treatment B and comparison 4. The study team assigned each beneficiary a matching weight to reflect the size of its matched set. All treatment beneficiaries received a weight of 1, whereas comparison beneficiaries received a weight equal to $\frac{n_j^T}{n_j^C}$, where n_j^T and n_j^C are the number of treatment and comparison beneficiaries in matched set j, respectively. In this example, comparisons 1–3 would each receive a weight of 0.333, and comparison 4 would receive a weight of 1. When matching with replacement, comparison beneficiaries can receive matching weights greater than 1; in a matched set containing treatment beneficiaries C and D and comparison beneficiary 5, comparison beneficiary 5 would receive a weight of 2.

After producing an initial match, the matching team checked balance using the diagnostic measures described in the introduction. If the initial balance did not meet the Centers for Medicare & Medicaid Services' standard of standardized differences within ± 0.1, the team refined the match by imposing constraints on particular variables. These constraints, called calipers, force or strongly encourage the matching algorithm to consider as potential matches only observations that meet certain criteria. For example, if balance was initially inadequate on a key covariate such as baseline Medicare expenditures, the team might implement a caliper stipulating that no matches could differ by more than $100 per beneficiary per month on this variable. Matching proceeded iteratively until the solution produced acceptable balance on all covariates.

For some awardees, the study team also incorporated constraints in the matching algorithm to ensure good balance separately within subgroups of interest, such as high-risk beneficiaries. Strategies for ensuring good subgroup balance included (1) exact-matching on the subgroup variable so that, for example, high-risk treatment beneficiaries can match only to high-risk potential comparison beneficiaries; (2) adding interaction terms to the propensity score model to allow the relationship between treatment status and the covariates to differ for different subgroups; and (3) in some cases, even matching the subgroups separately. When subgroup analyses were particularly important, the team checked balance separately by subgroup to determine whether the comparison group supported causal inference at this level.

2. Variations on the default strategy

Of course, with a set of awardees as diverse as those funded under HCIA R2, there were inevitably cases in which the default strategy was inappropriate. Deviations from the default strategy fell into four main categories: random assignment, provider-level matching, accounting for low participation rates, and propensity score weighting.

a. Random assignment

Two of the HCIA R2 awardees, the University of California at San Francisco (UCSF) and the University of Illinois at Chicago, randomly assigned beneficiaries to the treatment and control groups. For these awardees, the study team checked balance to verify that random assignment produced samples with similar distributions of observable covariates. This confirmation was particularly important for UCSF, which had only about 350 beneficiaries enrolled in the study, two-thirds of whom were assigned to the treatment group.

b. Provider-level matching

In one case, Altarum, the intervention occurred at the provider rather than the beneficiary level. Although the evaluation examined the intervention's effects on beneficiaries' outcomes, the study team matched at the provider level to mimic the treatment assignment mechanism. However, the team adjusted the process slightly to account for the fact that different providers had different numbers of attributed beneficiaries.

For Altarum, the team weighted the propensity score model by the number of beneficiaries attributed to the provider. Doing so allowed larger providers to influence the model more than smaller providers, as they would in a beneficiary-level analysis. When checking balance using provider-level data, the team also adjusted the matching weights to approximate the beneficiary-level balance. As when estimating the propensity score model, the team multiplied the matched set weights, calculated following the procedure described before, by weights representing the practice's size. With the additional practice size weights, the balance checks reflected the distribution of characteristics in the beneficiary sample.

c. Accounting for low participation rates

For other awardees, such as Mesa, the study team adopted an alternative propensity score modeling approach to account for low participation rates among those eligible to enroll in the intervention. An intent-to-treat considers all of those eligible to enroll as members of the treatment group, regardless of whether they ultimately participated in the intervention. However, if very few do participate, the characteristics of those eligible for treatment are less relevant to understanding the assignment mechanism than the characteristics of those who participate.

For this reason, rather than fitting a logistic regression to treatment and comparison observations in which the outcome variable is a treatment indicator, the team fit a logistic regression to members of the treatment group—all those eligible to participate in the intervention—for which the outcome variable is a participation indicator. This regression predicts the likelihood of participation, conditional on the covariates, among those eligible for the intervention. The study team then predicted propensity scores for the comparison group from the fitted regression model and used these propensity scores for matching. In line with the intent-to-treat design, the team matched all members of the treatment group regardless of whether they participated.

d. Propensity score weighting

The optimal matching algorithm is computationally intensive, especially for awardees with large treatment and potential comparison groups. For some awardees with exceptionally large samples, such as the Association of American Medical Colleges, the analysis used propensity score weighting instead of matching to expedite the process.

As for matching, the first step in propensity score weighting is to fit a propensity score model predicting the probability of assignment to the intervention conditional on the covariates. Rather than using the estimated propensity scores as input to an optimization procedure, however, propensity score weighting uses them directly to create matching weights. For HCIA R2, the study team created propensity score weights using the formula for estimating the average treatment effect on the treated, in line with the intent-to-treat strategy common to the awardees (Hirano and Imbens 2001):

$$(1) \qquad w_i = \begin{cases} 1 & \text{if } Z_i = 1 \\ \dfrac{\hat{p}_i}{1-\hat{p}_i} & \text{if } Z_i = 0 \end{cases}$$

In Equation (1), w_i is the matching weight assigned to beneficiary *i*, $\hat{p}_i$ is that beneficiary's estimated propensity score, and Z_i is the treatment indicator, with $Z_i = 1$ for treatment beneficiaries and $Z_i = 0$ for comparison beneficiaries. The team then normalized the weights to have the same sum in the treatment and comparison groups, for consistency with the matched set weights described previously.

C. Additional considerations

1. Outliers

To inform the matching approach, for each awardee the study team thoroughly explored the data. In many cases, this procedure revealed beneficiaries with extreme outlier values of key matching covariates, such as baseline Medicare expenditures or baseline hospitalizations. Across awardees, beneficiaries with extreme outlier values on these variables were much more likely to be potential comparison beneficiaries, although some awardees did have a few outliers in the treatment group. The study team's strategy for handling outliers depended on overall sample sizes and whether the outliers were in the treatment or potential comparison group. Including extreme outliers in the matching pool distorts the estimation of the propensity score model, so it is customary to exclude them at this stage if the potential comparison pool is sufficiently large. If the potential comparison pool was not large enough to exclude outliers, or if some treatment group members were outliers as well, the study team top-coded the variables with outliers at the 98th percentile, separately for treatment and potential comparison observations. That is, the team top-coded separately for treatment and potential comparison beneficiaries, and manually coded all values of the variable greater than the 98th percentile to the 98th percentile value. The team then fit the propensity score model and matched on the top-coded values, though in many cases it reported balance diagnostics on both the top-coded and original versions of the variable, for comparison.

2. Interventions with rolling enrollment

One challenge for evaluations using comparison samples is defining the point in time when the follow-up period for measuring outcomes should begin (and when the baseline period should end). Nearly all of the HCIA R2 awardees implemented interventions with rolling enrollment, whereby participants enrolled in the programs at different times rather than on a fixed intervention start date. For such interventions, participants' enrollment dates determine their baseline period, typically defined as the 12 to 24 months leading up to enrollment. However, because potential comparison observations by definition do not have an enrollment date, the study team could not define their baseline period in the same way. It is necessary therefore to assign them a pseudo-enrollment date.

In some cases—for example, when interventions for the treatment group begin with a seminal event observable in claims data, such as a hospital discharge or an emergency department visit—assigning a pseudo-enrollment date for potential comparison cases is straightforward. The date that they experienced the seminal event during the period of program operations defines the pseudo-enrollment date for each potential comparison case. However, assigning a pseudo-enrollment date to comparison cases is more difficult when a seminal event does not determine treatment group members' enrollment. In that case, rather than assigning each potential comparison beneficiary an arbitrary enrollment date, it is typical to consider several different pseudo-enrollment dates for each potential comparison beneficiary. For example, the study team might construct baseline covariates for each potential comparison observation beginning at one-

month intervals throughout a certain calendar year, so that the potential comparison pool contains 12 different versions of each potential comparison individual.

Considering several pseudo-enrollment dates for each potential comparison beneficiary minimizes the arbitrariness of baseline data construction, but it introduces challenges for matching because, without interference, traditional matching approaches are likely to produce matched samples that include more than one version of the same potential comparison beneficiary. So-called duplicate comparison beneficiaries complicate analysis because it is unclear how to model correlation between these repeated observations or how to represent duplication in the matching weights, among other difficulties. As a result, the study team generally preferred that the matched comparison group contain only one version of each comparison individual.

Historically, the only way to ensure uniqueness in the comparison group in this scenario was to undertake a time-consuming iterative deduplication process. However, for HCIA R2 the study team implemented a new approach called GroupMatch (Pimentel et al. 2019) that exploits the network properties of the optimal matching algorithm to integrate deduplication into the matching process. This approach streamlined matching for the many awardees that used rolling enrollment for which no seminal health event defined the enrollment date for program participants.

The GroupMatch approach ensures that no unique comparison beneficiary appears in the selected comparison group more than once, but does not guarantee that matched treatment and comparison beneficiaries have similar enrollment dates. To account for the effects of secular trends, the study team also sought to ensure a similar distribution of enrollment dates between the treatment and control groups. The team typically did so by exact-matching on the quarter of enrollment or by disallowing matches between treatment and comparison beneficiaries whose enrollment or pseudo-enrollment dates differed by more than a fixed amount, set at 30 to 90 days, depending on the awardee.

This page has been left blank for double-sided copying.

Appendix C:

Minimum detectable effects on expenditures and service use that program impact evaluations were powered to detect

This page has been left blank for double-sided copying.

To produce credible impact estimates, evaluations needed to have a large enough sample size to detect program effects less than or equal to 20 percent on total expenditures, number of hospitalizations, or number of ED visits. It is important to note that the minimum detectable effect had to be estimated in advance of a full data analysis, based on assumptions about the expected number of participants who could be identified in claims data and the variability in outcomes among these participants. In several cases, the usable treatment group sample was smaller and their outcomes more variable than the original assumptions projected. As a result, there were four awardees for which the impact evaluation was not able to detect (with 80 percent confidence) impacts of 20 percent on any of the core outcomes (Amerigroup, Montefiore, SCH, and UCSF).

Table C.1. Summary of minimum detectable effects (MDEs) that program evaluations were powered to identify, and the percent impact detected

	Percentage impact			Minimum detectable effect[a]		
Awardee	Total expenditures ($ PBPM)	Hospital admissions (per 1,000 beneficiaries)	ED visits (per 1,000 beneficiaries)	Total expenditures ($ PBPM)	Hospital admissions (per 1,000 beneficiaries)	ED visits (per 1,000 beneficiaries)
Medicare (N = 9)						
AAMC	< -1%	< -1%	< 1%	2.0%	3.3%	2.8%
Avera	-4%*	-1%	-9%***	5.5%	6.9%	6.6%
CCC	6%	2%	-21%***	21%	26%	20%
CCNC	< 1%	< 1%	< -1%	1.9%	2.4%	2.0%
CHIIC	-1%	-7%	-7%	17%	22%	17%
FSCL	10%*	NA	NA	15%	NA	NA
NMC	-11%	-1%	< -1%	17%	25%	20%
UCSF	-5%	-10%	-9%	30%	40%	35%
UKS	-4%	3%	-2%	15%	16%	10%
Medicaid (N = 8)						
Altarum	NA	NA	2%	NA	NA	2.7%
Amerigroup	NA	-19%	-13%	NA	75%	37%
Mesa[b]	NA	< 1%	11%***	NA	8.7%	6.9%
Montefiore	NA	-18%	-14%*	NA	34%	21%
NACHRI	NA	16%*	-10%***	NA	22%	9.1%
NYC H+H[b]	NA	-6%*	-7%**	NA	9.4%	8.2%
SCH	9%	34%	5%	62%	92%	43%
UIC	-21%**	-15%	-6%	26%	38%	14%
Medicare and Medicaid (N = 2)						
FPHNY	< -1%[c]	3%	4%	22%	18%	22%
VCC	-8%[c]	-2%	6%	23%	18%	13%

Table C.1 (*continued*)

[a] The minimum detectable effect is the size of the true impact for which the evaluation is 80 percent likely to find a statistically significant estimate in the study sample.

[b] The program enrolled both Medicare beneficiaries and Medicaid enrollees. Assessments of favorable impacts relied on eligible Medicaid enrollees only. For Mesa, this was because Medicaid enrollees accounted for 60 percent of the study sample and the participation rate was higher among this subgroup. For NYC H+H, this was because favorable program effects were limited to Medicaid enrollees eligible for the program during the first 9 months.

[c] Total expenditures data are available only for the Medicare beneficiaries.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; NA = not available; PBPM = per beneficiary per month.

Appendix D:

Methodology for Bayesian analysis

This page has been left blank for double-sided copying.

D.1. Introduction

In addition to the traditional frequentist analysis presented in the body of this report, the evaluation team estimated program impacts for each awardee using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value.

Drawing probabilistic conclusions requires external or prior evidence. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to the awardee being analyzed. This approach is an innovative advance on the typical way of specifying prior information for a Bayesian model, which conventionally relies on judgment and experience. Drawing prior information directly from related studies enables the Bayesian analysis to build on the current state of knowledge about interventions designed to slow Medicare and Medicaid cost increases and improve quality of care, in a data-driven way.

The following sections, first, describe the details of the methodological approach and, then, summarize the findings of the Bayesian analysis across awardees.

D.2. Approach

As in the primary analysis, the Bayesian analysis aimed to achieve both standardization across awardees and flexibility to tailor the approach to individual awardees. Balancing these two objectives with practical concerns, such as computational intensity, led the evaluation team to adopt a meta-regression approach. In a meta-regression, impact estimates from previous studies—in this case, the HCIA R1 data and the frequentist HCIA R2 estimates—serve as data. With this approach, the team fit a Bayesian model to the frequentist impact estimates for each HCIA R2 awardee, rather than to beneficiary-level data.

Meta-regression is an established technique for gaining the benefits of Bayesian modeling, such as probabilistic inference, without incurring high computational costs (Gelman et al. 2004). This analysis used the meta-regression approach to generate new impact estimates for HCIA R2 studies by combining the frequentist impact estimates with evidence from previous studies in a data-driven prior distribution.

The next sections describe first this innovative, data-driven prior and then the full regression model.

D.2.1. Prior distributions

Bayesian models draw on external, or prior, evidence to enhance precision and produce probabilistic inference. Conceptually, the prior distribution for the impact of an HCIA R2

awardee should represent the distribution of true impacts in a population of similar interventions for a similar target population (Statistical Modeling, Causal Inference, and Social Science n.d.).

For HCIA R2, the HCIA R1 awardees provide just such a population of programs. As a result, identifying the prior distribution need not be a theoretical exercise; the evaluation team could use the findings of HCIA R1 to develop an empirical prior. Specifically, they relied on the HCIA R1 meta-analysis data, which comprised awardee-specific impact estimates for each of the four core outcomes for the full post-intervention period, in addition to background characteristics on each awardee. Because variation in impacts over time is an important part of the HCIA R2 evaluation, the team supplemented the meta-analysis data with time-varying HCIA R1 impact estimates gathered from the publicly available HCIA R1 annual reports. In total, complete impact and background characteristics data were available for 87 awardees.

The evaluation team incorporated prior information from the HCIA R1 awardees into the analysis by fitting a meta-regression model to data sets that combine the impact estimates and standard errors from a single HCIA R2 awardee with similar information from all HCIA R1 awardees. They created separate data sets for each HCIA R2 awardee, so that, for example, the CHIIC analysis data combined CHIIC information with all of the HCIA R1 data, and the UCSF analysis data combined UCSF information with all of the HCIA R1 data. For awardees that served both Medicare and Medicaid populations, the team created a single data set combining Medicare and Medicaid impact estimates from the HCIA R2 awardee's frequentist analysis with all of the HCIA R1 data.

Importantly, meta-regression enabled the team to control for the background characteristics of the HCIA R1 awardees, so that the model could account for features associated with more or less favorable impacts when estimating a prior distribution for each HCIA R2 awardee. This strategy also implies that the prior for each HCIA R2 awardee derives primarily from the impacts of HCIA R1 awardees with similar background characteristics. In this way, the evaluation maintained a consistent modeling approach across awardees while tailoring the prior to each HCIA R2 awardee based on its characteristics.

D.2.2. Estimating impacts

D.2.2.1. Data

In a meta-regression, impact estimates from previous studies—in this case, the HCIA R1 data and the frequentist HCIA R2 estimates—serve as data. When combining estimates across outcomes and studies, as in this analysis, it is important to standardize the scale of the estimates. Standardization must account for different outcome measures, so that impacts on total Medicare expenditures, which are measured in dollars per beneficiary per month (PBPM), are commensurate with impacts on hospital admissions, which are measured in events per 1,000 beneficiaries.

Because the meta-regression includes estimates from different HCIA R1 and HCIA R2 awardees, the standardization approach must also account for the possibility that an impact of the same absolute magnitude would have a different interpretation in different contexts; a $5 reduction in expenditures would represent a much more substantial gain in a population with baseline expenditures of $100 PBPM than in a population with baseline expenditures of $1,000 PBPM. The evaluation team adopted the percentage impact scale as a way of standardizing estimates across outcomes and awardees because this framing accounts for awardee-specific context, is intuitive to health services researchers, and relies on information available in the HCIA R1 meta-analysis data.

For consistency with the frequentist analysis, for each HCIA R2 awardee the evaluation team analyzed the estimates presented in the body of the awardee-specific narrative; for example, if awardee A presented impact estimates for months 1–12 and 13–24 separately, and awardee B presented impact estimates for months 1–6, 7–12, 13–18, and 19–24 separately, the Bayesian analysis produced annual impact estimates for awardee A and semiannual estimates for awardee B.

To align the HCIA R2 analysis with information available from HCIA R1, the evaluation team produced Bayesian impact estimates for only the four core outcomes determined by the Centers for Medicare & Medicaid Services (CMS): total Medicare expenditures, hospital admissions, ED visits, and hospital readmissions. For a given HCIA R2 awardee, the analysis incorporated estimates presented in the awardee-specific narrative; for most awardees, this criterion excluded the readmissions outcome. For similar reasons, it typically excluded the Medicare expenditures outcome from the analysis of Medicaid-only awardees. Alignment with the HCIA R1 data also required the evaluation team to analyze full sample results only, because the HCIA R1 meta-analysis data did not include subgroup-specific impact estimates. The team violated this rule of thumb for several awardees for which the primary results were for subgroups of the full sample, enumerated in Table D.1. For these awardees, as for the awardees that served both Medicare and Medicaid populations, the team fit a single model to a data set that combined HCIA R1 information with information from each of the awardee subgroups.

Table D.1. Awardees without full sample results

Awardee	Subgroups included in Bayesian analysis
Avera	Long-stay SNF patients
	Short-stay SNF patients
University of Kansas	Critical access hospital patients
	Noncritical access hospital patients

SNF = skilled nursing facility.

For one awardee, UIC, the team fit the Bayesian model separately to frequentist impact estimates from the full sample and estimates from the high-risk subpopulation, to align with the approach preferred for the HCIA R2 meta-analysis.

D.2.2.2. Regression approach

After converting all impact estimates, and their standard errors, to the percentage impact scale, the evaluation team fit a Bayesian meta-regression according to the equation:

$$(1) \quad y_{ijt} = \alpha + \beta X_i + \gamma t + a_i + b_j + c_{p[i]} + \varepsilon_{ijt}$$
$$\varepsilon_{ijt} \sim N\left(0, \sigma_j^2 + s_{ijt}^2\right)$$

In Equation (1), y_{ijt} is the percentage impact for awardee *i* on outcome *j* in time period *t*. Each percentage impact in the data set corresponds to a single time period *t* for which the frequentist report presented impact estimates. For example, if the frequentist report presented annual estimates, the data set will include a y_{ijt} for each year of follow-up. However, for the HCIA R1 awardees, impacts were estimated for a range of different outcome intervals (for example, quarterly, semiannually, annually). Thus, a time indicator (*t*) was needed to account for these differences in measurement periods. The time period *t* was calculated as the midpoint of the follow-up interval, measured in months. The midpoint for the first year of follow-up for an awardee with impacts reported for 12-month periods would be Month 6, and the midpoint for the second six months of follow-up for an awardee with impacts estimated for semiannual intervals would be Month 9.

The team modeled y_{ijt} as an additive linear function of an overall intercept α, which represents the average impact across all outcomes and awardees, the effects of background characteristics X_i on impacts (Table D.2 provides a full listing), and the linear time effect γt. This linear time effect accounts for possible associations between the length of follow-up and the impacts; if impacts tend to be larger in later periods as programs learn the ropes, or conversely lower as enthusiasm wanes, this term will control for those relationships. The team also included three random effect terms: a_i is an awardee-specific effect, b_j is an outcome-specific effect, and $c_{p[i]}$ is a target population-specific effect, where $p[i]$ is the population that awardee *i* targeted. These random effect terms enabled the analysis to "borrow strength" across awardees, outcomes, and target populations, meaning that it could draw on awardees with more precise impacts to inform the estimates for awardees with less precise impacts. Importantly, however, the data dictated the extent to which the model could borrow strength across awardees, outcomes, and target populations; if the data indicated that impacts were very different for different awardees, the model would borrow less information along this dimension, whereas if the data indicated that impacts were similar across awardees, the model would borrow more information.

Finally, the error term includes two components. The first component, σ_j^2, represents variation in the true impacts across awardees and time for each outcome *j*. This component is thus outcome-specific, representing a belief that the amount of variation in true impacts could differ by outcome. The second component, s_{ijt}^2, is the standard error of the percentage impact y_{ijt} and thus

represents random error in the estimate. This component acts as a weight so that precise impact estimates contribute more to the model than imprecise impact estimates.

Following the current scholarly guidance,[10] the evaluation team used the following prior distributions:

- $\alpha \sim N(0, 0.05)$
- $\beta, \gamma \sim N(0,1)$
- $a_i \sim N(0, \sigma_a^2)$
- $b_j \sim N(0, \sigma_b^2)$
- $c_{p[i]} \sim N(0, \sigma_c^2)$
- $\sigma_j, \sigma_a, \sigma_b, \sigma_c \sim half\text{-}N(0,1)$

All priors are centered at zero, to avoid imposing assumptions about the sign of important quantities. Prior variances reflect the assumption that, in the absence of information otherwise, most terms are unlikely to be large on the scale of the data. For example, the half-normal prior on the standard deviation parameters indicates that (1) the standard deviations must be positive in sign and (2) standard deviations are unlikely to exceed 1, which corresponds to a percentage impact of 100 percent.

The prior distribution for the overall intercept term α is of particular interest because it represents the overall average impact across awardees, outcomes, and time periods. Although the standard in the literature is to use a standard normal prior, the evaluation team felt that the default prior lacked face validity for this analysis because on the percentage impact scale such a prior implies a 32 percent chance of seeing impacts larger in absolute value than 100 percent. Based on sensitivity tests comparing the effects of different priors, the team selected a prior standard deviation of 5 percent, implying that 32 percent of the true impacts are likely to fall outside ± 5 percent.

The evaluation team arrived at this regression equation after testing a series of alternatives that allowed for greater flexibility. For example, the team considered models in which the regression coefficients were outcome-specific, indicating that the relationship between a given control variable and the impacts could differ from outcome to outcome. However, they found that the quantity of interest—the probability of a favorable impact—was not sensitive to these alternatives, so they adopted the more parsimonious specification. In testing, the team found evidence to suggest that impacts on some outcomes were more variable than impacts on other outcomes, so they retained the outcome-specific signal variance term σ_j^2 rather than the more parsimonious σ^2.

[10] The developers of the Stan programming language, used for Bayesian estimation, publish a guide available at https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations.

Table D.2. Regression controls

Variable	Coding
Medicare	Yes/No
For-profit tax status	Yes/No
Academic affiliation	Yes/No
Used health information technology	Yes/No
Provided behavioral health services	Yes/No
Provided telemedicine	Yes/No
Used community health workers	Yes/No
Number of intervention sites	Integer
Rural location	Yes/No
Received no-cost extension	Yes/No
Barriers to enrollment	Yes/No
Staff turnover challenges	Yes/No
Implementation effectiveness	Effective, partly effective, or ineffective
Number of enrolled participants	Integer
Provided direct services	Yes/No
Targeted high-risk population	Yes/No
Number of post-intervention quarters	Integer

D.2.2.3. Inference

The evaluation team fit the regression model above to separate data sets for each HCIA R2 awardee. Each data set contained the HCIA R2 awardee's percentage impacts for each core outcome and time period presented in the awardee-specific narrative, as well as the percentage impacts for the HCIA R1 awardees for all core outcomes and the most granular time periods available. From the model fit the team derived three types of results: (1) estimates of impacts and their uncertainty on the percentage impact and original scales, (2) probabilities of achieving favorable impacts, and (3) implied prior distributions.

The evaluation team obtained awardee-specific impact estimates from the model fit by summarizing the posterior distribution of the impact, given by the linear predictor in the regression equation, for each combination of outcome j and time period t included in the main body of the awardee-specific narrative for the HCIA R2 awardee. These estimates were on the percentage impact scale, which the team converted to the original scale of the data by multiplying through by the counterfactual mean for that awardee. The counterfactual mean for each awardee, representing the outcomes that would have been observed in the intervention group without the intervention, is equal to the mean of the outcome variable in the intervention group in the post-intervention period minus the estimated treatment effect from the frequentist analysis.

The team obtained probabilities from the posterior distribution of the impact by comparing the distribution to substantively important threshold values, such as favorable impacts of 1, 5, and 10 percent. For example, the proportion of the distribution that lies below -0.01 is the probability of a favorable impact of 1 percent or more.

From the model fit they also calculated the prior distribution for each HCIA R2 awardee's impact on each outcome in each time point according to the equation:

$$(2) \quad \tilde{y}_{ijt} \sim N\left(\alpha + \beta X_i + \gamma t + b_j + c_{p[i]}, \sigma_a^2 + \sigma_j^2\right)$$

Because, before observing the HCIA R2 awardee's data, that awardee's awardee-specific effect a_i is unknown, the equation for the prior excludes this term from the mean of the distribution and instead accounts for uncertainty in the awardee-specific effect in the distribution's variance. The team calculated the prior mean for each awardee as the sum of the estimated coefficients in the linear predictor, that is, $\hat{\tilde{y}}_{ijt} = \hat{\alpha} + \hat{\beta} X_i + \hat{\gamma} t + \hat{b}_j + \hat{c}_{p[i]}$, and likewise calculated the prior variance as the sum of the estimated variance components, $Var\left(\hat{\tilde{y}}_{ijt}\right) = \hat{\sigma}_a^2 + \hat{\sigma}_j^2$. To obtain prior probabilities, they computed the tail probabilities of the normal distribution with mean $\hat{\tilde{y}}_{ijt}$ and variance $Var\left(\hat{\tilde{y}}_{ijt}\right)$.

For awardees that sought to engage both Medicaid and Medicare populations, the evaluation also generated pooled results for outcomes common to both populations by weighting the posterior distributions for Medicaid and Medicare impacts by the percentage of enrollees in each population. For example, if awardee A had 30 Medicaid enrollees and 70 Medicare enrollees, then the pooled impact estimates would be the sum of 30 percent of the Medicaid impact estimates and 70 percent of the Medicare impact estimates. The impact estimates and posterior and prior probabilities were then calculated from the pooled distributions.

D.3. Results

This section discusses several facets of the Bayesian analysis results. First, it investigates features of the meta-regression—control variables and structural components—that influence the results. Then it summarizes the results of the innovative empirical prior development strategy and compares the model's empirical priors to the impact results.

D.3.1. Main meta-regression influences

Before combining the HCIA R1 data with HCIA R2 data, the evaluation team fit the meta-regression model to the HCIA R1 data only. Doing so was both an important way of assessing its appropriateness for the problem and of identifying features that drive impacts in the HCIA R1

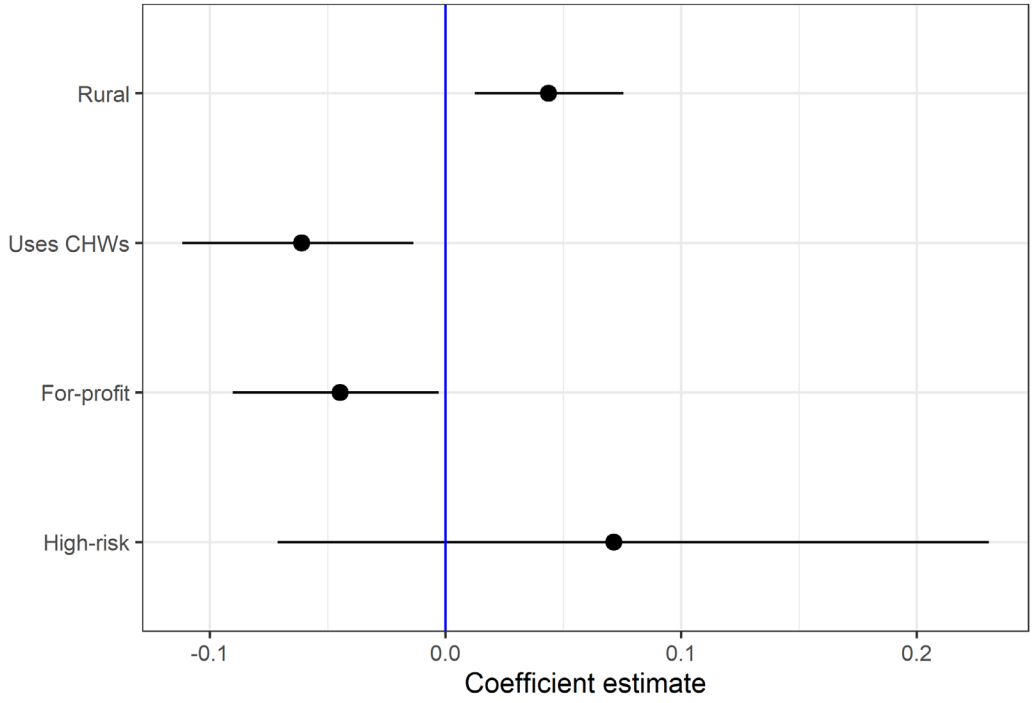
data. This section discusses some of the key results from the analysis of HCIA R1 data, which in turn informed the team's understanding of the HCIA R2 impact results.

D.3.1.1. Influential control variables

Several control variables were strongly associated with impacts in the meta-regression of HCIA R1 data, as Figure D.1 shows. In the figure, the x-axis is on the percentage impact scale; because reductions in outcomes are favorable for CMS's four core outcomes, a negative point estimate indicates an association with favorable impacts and a positive point estimate indicates an association with unfavorable impacts.

Two variables, use of community health workers and for-profit tax status, were strongly associated with favorable impacts in the meta-regression of HCIA R1 awardees. Another two variables, rural location and high-risk target population, were strongly associated with unfavorable impacts. These results largely corroborate the findings of the HCIA R1 meta-analysis, which focused on impacts on total Medicare expenditures only.

Figure D.1. Control variables strongly associated with impacts



Source: Mathematica's analysis of HCIA R1 data.

Note: Points represent coefficient estimates and horizontal lines mark the 95 percent credible interval for the estimate, calculated from the 2.5 and 97.5 quantiles of the posterior distribution for the parameter. The blue vertical line marks zero to aid in determining whether the uncertainty intervals contain this value.

CHW = community health worker; HCIA R1 = Round 1 of the Health Care Innovation Awards.

No other control variable was strongly associated with impacts, so these results suggest that all else equal, the model will anticipate less favorable impacts for HCIA R2 awardees located in

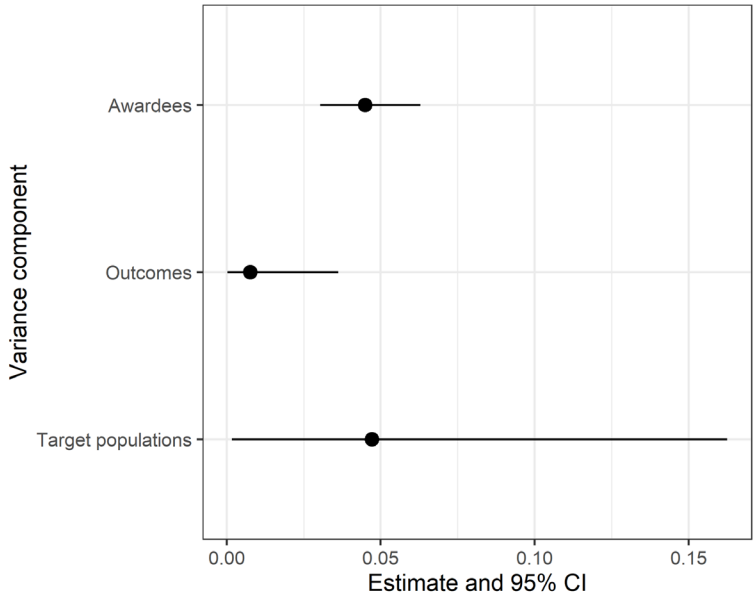
rural areas or targeting high-risk populations and will anticipate more favorable impacts for for-profit awardees or awardees that use community health workers. Although three of these associations are consistent with other literature, one—adverse results for programs targeting high-risk populations—is not. Nearly all studies of care coordination interventions that had random assignment designs find that favorable impacts, if any, tend to be greatest for the higher-risk subset of patients (Brown et al. 2012). In this context, the wide uncertainty interval around the estimate suggests that this unexpected association is far from consistent across studies.

D.3.1.2. Influential components

In addition to determining whether certain characteristics drive favorable or unfavorable prior probabilities for the HCIA R2 awardees, the results of the prior analysis show whether certain components of the model are more influential than others.[11] After controlling for background characteristics, the data vary along three dimensions: awardees, outcomes, and target populations. Variation across awardees describes how the average impact across outcomes differs from one awardee to another; variation across outcomes describes how impacts on different outcomes vary within a single awardee; and variation across target populations describes how the average impact differs across groups of awardees seeking to engage different populations.

The evaluation team estimated each of these variance components as part of the meta-regression model. Comparing the estimates of the components helps to determine which are more influential; Figure D.2 depicts the estimated variance components and their uncertainty intervals.

Figure D.2. Meta-regression variance components



CI = credible interval.

[11] As an example of this type of analysis, see Gelman (2005).

Figure D.2 shows that, in the HCIA R1 data, there is much more variation in impacts across awardees and target populations than there is across outcomes for the same awardee. In practice, the prominence of variation across awardees over variation across outcomes leads to similarity across outcomes in prior probabilities and percentage impacts from the Bayesian model; put differently, the model borrows more strength across outcomes *within* an awardee than *across* awardees. For example, for an awardee for which the frequentist impact estimates on the four core outcomes ranged from -22 to -9 percent, the Bayesian impact estimates ranged only from -10 to -9 percent, reflecting the pattern observed in the HCIA R1 data. Although variance across target populations also appears to be an important driver in the HCIA R1 data, its uncertainty interval is very large, so this component does not exert as much influence over the results as its point estimate suggests.

D.3.2. Impact results

The ability to draw probabilistic conclusions is a primary advantage of the Bayesian approach. For this reason, this appendix presents the results of the Bayesian impact analysis on the probability scale. Table D.3 summarizes the main impact results for each awardee, giving the probability of a favorable impact of 1 percent or more on each of CMS's four core outcomes, as applicable, for the time period and sample included in the HCIA R2 impact assessment (Chapter III). Table D.3 shows that, although impact probabilities tend to be consistent across outcomes for the same awardee, they vary greatly across awardees, from roughly 10 percent to over 90 percent.

Table D.3. Probability of a favorable impact of 1 percent or more

Awardee	Total expenditures	Hospital admissions	ED visits	Readmissions
Medicare (N = 9)				
AAMC	73%	67%	76%	n.a.
Avera[a]	93%	90%	95%	91%
CCC	41%	38%	44%	39%
CCNC	9%	8%	14%	n.a.
CHIIC	31%	28%	33%	n.a.
FSCL	3%	n.a.	n.a.	n.a.
NMC	83%	81%	84%	82%
UCSF	52%	50%	54%	n.a.
UKS	20%	18%	22%	n.a.
Medicaid (N = 7)[b]				
Amerigroup	n.a.	91%	93%	n.a.
Mesa	n.a.	10%	12%	10%
Montefiore	n.a.	79%	81%	n.a.

Table D.3 (*continued*)

Awardee	Total expenditures	Hospital admissions	ED visits	Readmissions
NACHRI	n.a.	26%	31%	n.a.
NYC H+H[c]	n.a.	25%	30%	n.a.
SCH	79%	77%	81%	n.a.
UIC[a]	97%	97%	98%	n.a.
Medicare and Medicaid (N = 2)				
FPHNY	28%	21%	26%	22%
VCC	77%	58%	63%	60%

Source: Mathematica's analysis of HCIA R2 and HCIA R1 data.

Notes: For each awardee, this table gives the probability of a favorable impact of 1 percent or more on each of CMS's four core outcomes over the follow-up period and sample included in the meta-analysis.

[a] The cumulative probability over the full 24-month follow-up period was not available. The reported probability is for the 13 to 24-month follow-up period.

[b] No Bayesian analysis was conducted for Altarum because the awardee did not anticipate that the intervention would have an effect on the core outcomes.

[c] The cumulative probability over the Medicaid enrollees eligible for the program during the first 9 months was not available. The reported probability is based on the full Medicaid sample.

n.a. indicates that the awardee's narrative did not present impact estimates for this outcome variable; few Medicaid awardees presented impacts on expenditures. This table includes impacts on readmissions for some awardees because the Bayesian analysis jointly models impacts on all four of CMS's core outcomes where they are available.

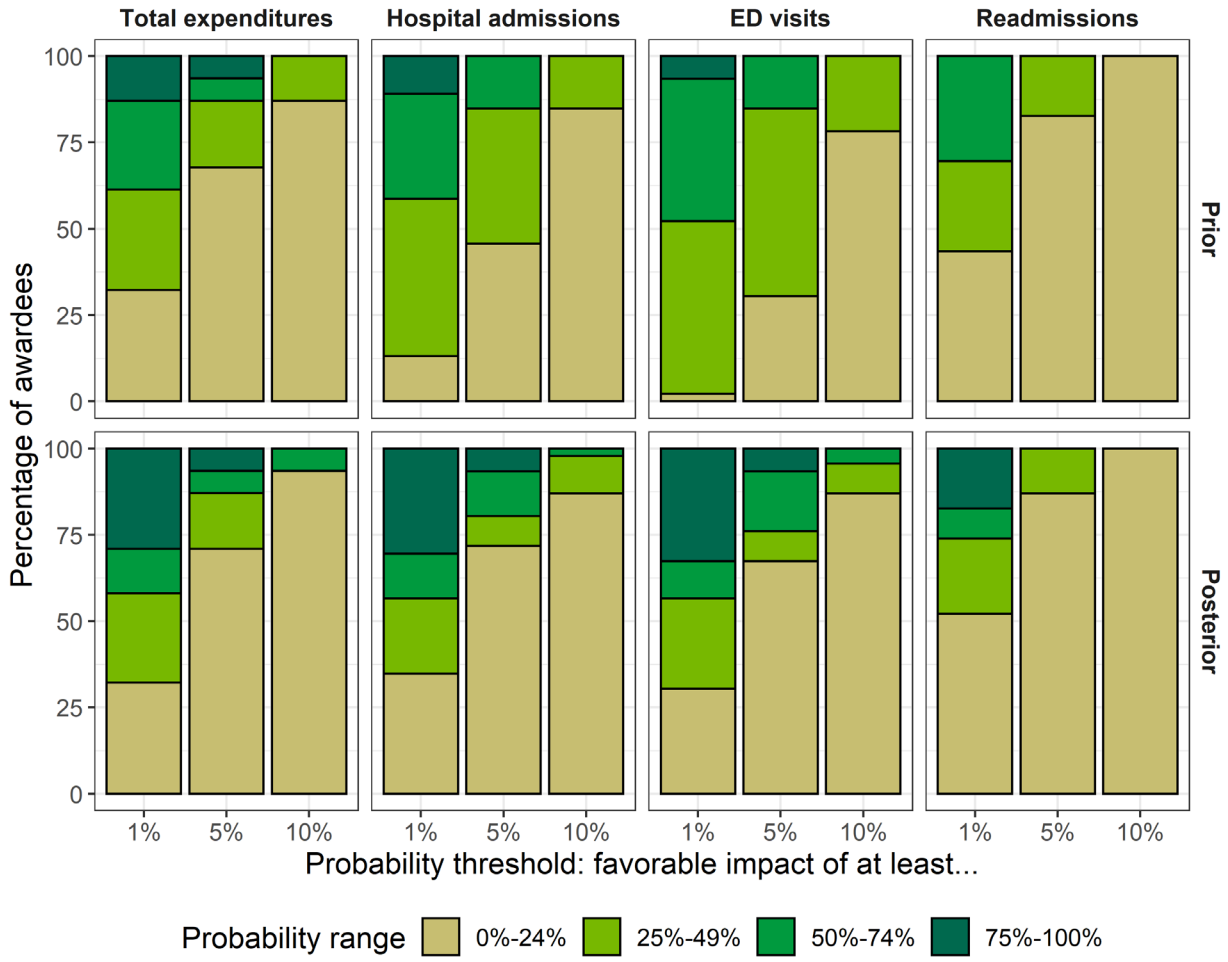
Alongside impact probabilities, this appendix also presents comparable probabilities summarizing the data-driven prior distributions estimated as part of the analysis. The evaluation team's innovative approach to developing empirical prior distributions, described earlier, created a tailored prior distribution for each combination of awardee, outcome, and follow-up period. The variation in these priors is itself an interesting result of the study because it reflects the association between an awardee's characteristics and impacts in the HCIA R1 data. The evaluation team explored the sources of this variation before summarizing impact results across HCIA R2 awardees.

It is especially instructive to compare prior probabilities—representing the model's prediction for the impact estimates for a given HCIA R2 awardee before looking at the frequentist impact estimates for that awardee—to posterior probabilities, which represent a complete understanding of an HCIA R2 awardee's impacts after combining frequentist estimates of the awardee's impacts with the prior evidence from HCIA R1. This comparison in some sense conveys new insight gained from the HCIA R2 data, or how the HCIA R2 data differed from the model-based predictions.

To summarize concisely across many awardees and time points, the evaluation team calculated the percentage of awardees with an impact probability in each quarter of the probability range: 0–24, 25–49, 50–74, and 75–100 percent. Figure D.3 shows how awardees' prior and posterior impact probabilities are distributed across these ranges at three different probability thresholds:

favorable impacts of 1, 5, and 10 percent. For example, the dark green bars in the left-hand panel represent the proportion of awardees with at least a 75 percent probability of a favorable impact of 1, 5, or 10 percent based on the HCIA R1 data alone (top) and HCIA R1 and R2 data combined (bottom).

Figure D.3. Prior and posterior probabilities of favorable impacts across outcomes



Sources: Mathematica's analysis of frequentist impact estimates from awardees' narratives. The Bayesian analysis also incorporated HCIA R1 data.

Note: Rows in the plot represent phases of the analysis; the top row shows results from the prior distribution, representing the model's predictions before seeing the frequentist impact estimates for the HCIA R2 awardees. The bottom row shows results from the posterior distribution—that is, the Bayesian impact estimates—representing our understanding of the impacts after combining information from the prior with the frequentist impact estimates. Each column shows one of CMS's four core outcomes. In each panel, the x-axis marks three probability thresholds: favorable impacts of at least 1, 5, or 10 percent. The bars are divided into segments based on the proportion of awardees with impact probabilities in each quarter of the probability range (0–100 percent).

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards.

A single panel of the plot shows how impact probabilities at different levels are concentrated across the 0 to 100 percent spectrum. For example, the bottom-left panel, which summarizes the posterior probabilities of impacts on total Medicare expenditures, shows that across HCIA R2 awardees, there are generally low probabilities of large impacts on Medicare expenditures: at the 5 and 10 percent levels, the yellow segments, representing a posterior probability of less than 25 percent, account for 75 to 90 percent of awardees. However, at the 1 percent level awardees are divided more evenly among the probability quartiles, with about 25 percent of awardees in both the bottom (less than 25 percent probability of an impact of 1 percent or more) and top (at least a 75 percent probability of an impact of 1 percent or more) quarters.

Comparing the bottom-left and top-left panels shows how incorporating the HCIA R2 frequentist impact estimates changed the distribution of impacts on total Medicare expenditures. The proportion of awardees with prior (top) and posterior (bottom) probabilities below 25 percent is similar at each probability threshold, suggesting that similar proportions of HCIA R2 and HCIA R1 awardees had null effects, conditional on their background characteristics. However, the posterior probabilities panel indicates that a higher proportion of HCIA R2 awardees had high probabilities of favorable impacts of 1 percent or more than anticipated based on the HCIA R1 data; in the bottom panel about 25 percent of the bar is dark green, indicating an impact probability of 75 percent or more, compared to only 10 percent in the top panel.

A similar pattern emerges across outcomes. In general, the prior (top row) anticipates that most awardees will have probabilities of less than 75 percent of favorable impacts of 1 percent or more, but the posterior (bottom row) shows that awardees' probabilities are more evenly distributed across the probability quartiles. In particular, a higher proportion of awardees had a probability of a favorable impact of 1 percent or more that exceeded 75 percent, counter to prior expectations. In practical terms, these results indicate that for most awardees, there was a low probability of substantial impact on each outcome, in agreement with the prior. However, more awardees than expected had a high (75 percent or greater) probability of a favorable impact of 1 percent or more. Importantly, this pattern is not as marked for the higher probability thresholds of 5 or 10 percent or more; taken together, these results point to a high probability of small impacts but a low probability of more substantial impacts.

This page has been left blank for double-sided copying.

Appendix E:

Model features used to assess evidence of favorable impacts

This page has been left blank for double-sided copying.

Table E.1. Features of models used to determine whether programs had evidence of favorable impact on core outcomes

Awardee	Type of enrollment[a]	Type of participant[b]	Total number of beneficiaries served	Percentage of eligible beneficiaries who participated	Number of treatment group beneficiaries in impact analysis	Evaluation design	Impact estimate based on participants or eligible beneficiaries	Subgroups evaluated in impact analysis	Follow-up periods used in impact analysis
Medicare									
AAMC	Passive	Indirect	128,721	100%	145,938[c]	DD[d]	Participants only	• Full sample • Full sample excluding University of Iowa • Program maturity groups[e]	• 1–3 months • 4–6 months
Avera	Passive	Direct and indirect	11,192	100%	9,608	DD	Participants only	• Long-term NF residents • SNF residents	• 1–12 months • 13–24 months • 1–24 months
CCC	Passive	Direct	8,016	Unknown	900	CS	Participants only	• Full sample	• 1–12 months • 13–24 months
CCNC	Passive and active	Direct	328,806	55%	110,968	DD	All eligible beneficiaries	• Full sample • Program maturity groups[e]	• 1–6 months • 7–12 months • 1–12 months • 13–18 months • 19–24 months • 13–24 months • 1–24 months
CHIIC	Active	Direct and indirect	6,489	23%	1,924	DD	Participants only	• Full sample • Program maturity groups[e]	• 1–12 months • 13–24 months
FSCL	Passive	Indirect	5,803	38%	2,097	DD	All eligible beneficiaries	• Full sample	• 1–12 months

Table E.1 (*continued*)

Awardee	Type of enrollment[a]	Type of participant[b]	Total number of beneficiaries served	Percentage of eligible beneficiaries who participated	Number of treatment group beneficiaries in impact analysis	Evaluation design	Impact estimate based on participants or eligible beneficiaries	Subgroups evaluated in impact analysis	Follow-up periods used in impact analysis
NMC	Active	Direct	1,903	8%	430	CS	Participants only	• Full sample • Controlling for mortality	• 1–6 months • 7–12 months • 1–12 months
UCSF	Active	Direct	512	100%	358	RCT	Participants only	• Full sample	• 1–6 months • 7–12 months • 1–12 months
UKS	Passive and active	Indirect	7,334	100%	920[f]	CS	Participants only	• AMI intervention • Stroke intervention	• 1–6 months • 1–12 months • 12–24 months
Medicaid									
Altarum	Passive	Indirect	949,164	100%	94,944[g]	DD	Participants only	• Full sample (beneficiaries) • Full sample (providers) • Program maturity groups[e]	• 1–12 months • 13–24 months • 1–24 months
Amerigroup	Active	Indirect	860	30%	299	DD	Participants only	• Full sample • Female participants only	• 1–6 months • 7–12 months • 1–12 months • 13–18 months • 19–24 months • 13–24 months
Montefiore	Passive	Direct and indirect	6,559	Unknown[h]	2,069[f]	DD	Participants only	• Full sample • Depression severity groups	• 1–12 months • 13–24 months • 1–24 months

Table E.1 (*continued*)

Awardee	Type of enrollment[a]	Type of participant[b]	Total number of beneficiaries served	Percentage of eligible beneficiaries who participated	Number of treatment group beneficiaries in impact analysis	Evaluation design	Impact estimate based on participants or eligible beneficiaries	Subgroups evaluated in impact analysis	Follow-up periods used in impact analysis
NACHRI	Passive	Indirect	8,111	100%	3,836[f]	DD	Participants only	• Full sample • Health status using 3M CRG • By site	• 1–12 months • 13–24 months • 1–24 months
SCH	Passive	Direct	813	17%	516	DD	Participants only	• Full sample	• 1–12 months • 13–24 months • 1–24 months
UIC	Passive	Direct	8,455	20%	3,131[f]	RCT	Participants only	• Full sample • Risk of unnecessary health care service use	• 1–12 months • 13–24 months • 1–24 months
Medicare and Medicaid									
FPHNY	Active	Direct	2,775	3%	1,310 Medicaid 327 Medicare	DD	Participants only	• Full sample • Medicare only • Medicaid only	• 1–12 months • 13–24 months • 25–36 months • 1–36 months
Mesa	Passive	Direct	12,431	75% Medicaid 44% Medicare	2,872 Medicaid 1,750 Medicare	DD	All eligible beneficiaries	• Medicare only • Medicaid only	• 1–3 months • 4–6 months • 7–9 months • 10–12 months • 1–12 months
NYC H+H	Active	Direct	83,946	15% Medicaid 18% Medicare	9,747 Medicaid 9,134 Medicare	DD	All eligible beneficiaries	• Full sample • Medicare only • Medicaid only • Program maturity groups[e]	• 1–6 months • 7–12 months • 1–12 months

Table E.1 (*continued*)

Awardee	Type of enrollment[a]	Type of participant[b]	Total number of beneficiaries served	Percentage of eligible beneficiaries who participated	Number of treatment group beneficiaries in impact analysis	Evaluation design	Impact estimate based on participants or eligible beneficiaries	Subgroups evaluated in impact analysis	Follow-up periods used in impact analysis
VCC	Active	Direct	4,366	2%	2,532 Medicaid 420 Medicare	DD	Participants only	• Full sample • Medicare only • Medicaid only	• 1–12 months • 13–24 months • 25–36 months • 1–36 months

[a] Some programs required informed consent or some other form of direct contact with potential participants triggered by a predefined event (active enrollment), whereas others served everyone who sought care at a participating site who met the program eligibility criteria (passive enrollment). A few programs had multiple components, some of which involved passive enrollment and others that involved active enrollment.

[b] A direct participant is an individual who receives care or services paid for by HCIA R2 program funding, such as care coordination services. An indirect participant is anyone who does not receive such services, but who benefits from the HCIA R2 funding nonetheless. For these participants, HCIA R2 funding generally helps to service providers, such as by funding to hire program staff, train intervention staff, and purchase or develop technology. These resources, in turn, can enhance and support clinicians' ability to deliver high quality, cost-efficient care to participants. A few programs had multiple components, some of which required direct participation and some for which participation was indirect. For example, CHIIC included both care coordination services, which required direct participation, and organizational-level quality improvement activities. The benefit to participants resulting from these activities was indirect.

[c] For AAMC, the number of treatment group beneficiaries included in the impact analysis exceeded the awardee's count of the total number of beneficiaries served by the program because the attribution rules used in the impact analysis (which relied on claims-based program enrollment criteria) differed from the reporting process used by the participating academic medical centers. In addition, the impact analysis used several methods to generate lists of providers participating in the program and chose the approach that most closely aligned the reports provided by the awardee, but the final list was not an exact match to the awardee's report.

[d] Impact evaluation relied on a repeat cross-sectional difference-in-differences design. All other difference-in-differences models relied on longitudinal cross-sectional designs.

[e] The impact evaluation examined whether program effects on the core outcomes differed for beneficiaries enrolling during the first nine months of implementation compared to impacts for those enrolling during the later months, after the programs had time to mature.

[f] The number of treatment group beneficiaries included in the impact analysis and reported in this table differs from the number of treatment group beneficiaries reported in Chapter III, Table III.1, because this table reports the total number of treatment group beneficiaries whereas Table III.1 reports only the number of treatment group beneficiaries included in the analysis that aligns with the features described in the table. For UKS, Table III.1 reports only the number of AMI participants. For Montefiore and NACHRI, Table III.1 reports only the number of participants who were followed for 24 months. For UIC, Table III.1 reports only the number of participants at higher risk of unnecessary health care service use.

[g] The impact analysis for Altarum examined only the training and TA component of the intervention. Therefore, the number of treatment group beneficiaries reported reflects the number of beneficiaries in the analysis who participated in this intervention component. The total number of beneficiaries served by this intervention component was 157,985.

[h] The participation rate could not be estimated because the program's most important eligibility criteria are not observed in the claims data.

AMI = acute myocardial infarction; CAH = critical access hospital; CRG = clinical risk groups; CS = cross-sectional; DD = difference in differences; HCIA R2 = Round 2 of the Health Care Innovation Awards; NF = nursing facility; RCT = randomized controlled trial; SNF = skilled nursing facility; TA = technical assistance.

Appendix F:

Rationale for selection of study sample and follow-up period for awardees with impact analyses

This page has been left blank for double-sided copying.

Table F.1. Rationale for selection of study sample and follow-up period for awardee models with impact analyses

Awardee	Participants or eligible beneficiaries	Full sample or subgroup	Follow-up period	Evidence of favorable effects on core outcomes?
Medicare				
AAMC	The intervention affected all eligible beneficiaries, so the participant group was the same as the eligible group.	The full sample was used because there were no discernible impacts among relevant subgroups.	The evaluation used a1- to 3-month follow-up period because the awardee expected the intervention to have short-term impacts.	No; estimated effects were small and unfavorable for ED visits, and not statistically significant.
Avera	The intervention affected all eligible beneficiaries, so the participant group was the same as the eligible group.	The sample included only long-term skilled nursing facility residents because program effects were limited to this subgroup.	The evaluation used cumulative effects because results were similar across follow-up periods.	Yes; effects were consistently favorable and statistically significant for total expenditures and ED visits.
CCC	The intervention affected all eligible beneficiaries, so the participant group was the same as the eligible group.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 1- to 12-month follow-up period because the awardee expected the program to have its most significant impact during the first 90 days of enrollment.	No; estimated effects on hospitalizations and total expenditures were not favorable.
CCNC	The evaluation used all eligible beneficiaries to minimize risk of selection bias. The participation rate was 48 percent.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used cumulative effects because results were similar across follow-up periods.	No; estimated effects on hospitalizations and total expenditures were not favorable.
CHIIC	The evaluation used participants only because the program had a low (11 percent) participation rate.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	1The evaluation used a 3- to 24-month follow-up period because the awardee expected it would take time for the intervention to achieve impacts on core outcomes.	No; estimated effects were small and not statistically significant.
FSCL	The evaluation used all eligible beneficiaries to minimize risk of selection bias. The participation rate was 38 percent.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 1- to 12-month follow-up period because the awardee expected the intervention to have short-term impacts.	No; estimated effects on total expenditures were not favorable.
NMC	The evaluation used participants only because the program had a low (8 percent) participation rate.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 1- to 12-month follow-up period because the awardee expected the intervention to have short-term impacts.	No; estimated effects were small and not statistically significant.

Table F.1 (*continued*)

Awardee	Participants or eligible beneficiaries	Full sample or subgroup	Follow-up period	Evidence of favorable effects on core outcomes?
UCSF	The evaluation used all beneficiaries randomized into a treatment group of the RCT.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 1- to 12-month follow-up period because only a small number of participants could be evaluated over 24 months.	No; estimated effects were moderate and in a favorable direction but were too imprecise to support a favorable assessment.
UKS	The intervention affected all eligible beneficiaries, so the participant group was the same as the eligible group.	The evaluation used patients with AMI treated at CAHs. The program also enrolled patients who experienced a stroke, but the sample size was too small to detect impacts of reasonable size.	The evaluation used a 1- to 12-month follow-up period because the awardee expected that the program would have its greatest impact during the AMI or stroke episode at the time of enrollment.	No; estimated effects were small and not statistically significant.
Medicaid				
Altarum	The intervention affected all eligible beneficiaries, so the participant group was the same as the eligible group.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used cumulative effects because results were similar across follow-up periods.	No; the core outcomes were not relevant to the intervention. The impact on dental-related ED visits was not favorable.
Amerigroup	The evaluation used participants because of the small sample size.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 1- to 12-month follow-up period because the median length of enrollment in the program was 3 months and the awardee expected the intervention to have short-term impacts.	No; estimated effects were moderate and in a favorable direction but not statistically significant. In addition, in Year 2, the estimated effects were in an unfavorable direction.
Mesa[a]	The evaluation used all eligible beneficiaries to minimize risk of selection bias. The participation rate was 75 percent.	The evaluation used the Medicaid sample because they were more likely to receive mobile emergency services compared to Medicare beneficiaries.	The evaluation used a 1- to 12-month follow-up period because the awardee expected the intervention to have short-term impacts.	No; estimated effects were not favorable.
Montefiore	The evaluation used participants because most beneficiaries who screened positive for depression enrolled in the intervention.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 13- to 24-month follow-up period because the awardee expected it would take time for the intervention to achieve impacts on core outcomes.	Yes; effects were consistently favorable and statistically significant for ED visits. Estimated effects on hospitalizations were moderate but not statistically significant.

Table F.1 (*continued*)

Awardee	Participants or eligible beneficiaries	Full sample or subgroup	Follow-up period	Evidence of favorable effects on core outcomes?
NACHRI	The intervention affected all eligible beneficiaries, so the participant group was the same as the eligible group.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 13- to 24-month follow-up period because the awardee expected it would take time for the intervention to achieve impacts on core outcomes.	No; estimated effects on core outcomes were not consistently favorable.
NYC H+H[a]	The evaluation used all eligible beneficiaries because it was not possible to replicate enrollment criteria in claims data.	The evaluation used Medicaid beneficiaries enrolled during the first 9 months because they were in poorer health and more frequent users of the ED before enrollment.	The evaluation used a 1- to 12-month follow-up period because the awardee expected the intervention to have short-term impacts.	Yes; estimated effects on hospitalizations and ED visits were favorable and statistically significant. Expenditure data were not available.
SCH	The evaluation used participants because of the small sample size.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used a 13- to 24-month follow-up period because the awardee expected it would take time for the intervention to achieve impacts on core outcomes.	No; estimated effects on hospitalizations and ED visits were not favorable.
UIC	The evaluation used all beneficiaries randomized into the treatment group of the RCT.	The evaluation used higher-risk children because the intervention effects were limited to this subgroup.	The evaluation used cumulative effects because results were similar across follow-up periods.	Yes; effects on total expenditures were favorable and statistically significant. Estimated effects on hospitalizations and ED visits were also favorable but not statistically significant.
Medicare and Medicaid				
FPHNY	The evaluation used participants because the program had a low (2 percent) participation rate.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used cumulative effects because results were similar across follow-up periods.	No; estimated effects were not statistically significant and not consistently favorable.
VCC	The evaluation used participants because the program had a low (4 percent) participation rate.	The evaluation used the full sample because there were no discernible impacts among relevant subgroups.	The evaluation used cumulative effects because results were similar across follow-up periods.	No; effect sizes were moderate but not statistically significant and not consistently favorable.

[a] The program enrolled both Medicare beneficiaries and Medicaid enrollees. Assessments of favorable impacts relied on eligible Medicaid enrollees only. For Mesa, this was because Medicaid enrollees accounted for 60 percent of the study sample and the participation rate was higher among this subgroup. For NYC H+H, this was because favorable program effects were limited to Medicaid enrollees eligible for the program during the first nine months.

AMI = acute myocardial infarction; CAH = critical access hospital; ED = emergency department; RCT = randomized controlled trial.

This page has been left blank for double-sided copying.

Appendix G:

List of variables used to assess association between program characteristics and program impacts

This page has been left blank for double-sided copying.

Table G.1. List of characteristics used to identify factors associated with evidence of favorable impacts

Program component	Number of awardees with feature
Behavioral health was a principal intervention component[a]	3
Telehealth was a principal intervention component[a,b]	6
Health IT was a principal intervention component[a]	6
Innovation had a care management component	7
Program or awardee characteristics	
Served a predominantly socially fragile population[a,c]	8
Program leadership was hospital based[a]	7
Program was community based (rather than provider based)	4
Program focused on individual patient care (rather than transforming provider practice)[a]	8
Awardee had previous experience implementing similar programs[a]	8
Awardee used active (as opposed to passive) enrollment[d]	6
Awardee used clinical judgement or factors not available in claims to identify eligible participants	5
Awardee weakened, loosened, or changed eligibility criteria during program implementation	4
Non-clinical staff were used as frontline intervention service providers[a]	8
Implementation experience	
Awardee was effective in enrolling beneficiaries	9
Awardee was effective in delivering services	11
Program services were delivered at a somewhat high or high level of quality and intensity	11
Awardee was effective in hiring, retaining, and training staff	11
Awardee was effective in engaging program participants	11
Awardee was effective in engaging clinicians and care givers in the program	7
Management staff turnover was a small, moderate, or major challenge	6
Frontline staff turnover was a moderate or major challenge	7
Intervention had robust partnership participation	4
Intervention was guided by strong awardee leadership	9

Source: Mathematica's implementation evaluation of the HCIA R2 awardees.

[a] Indicates variables strongly associated with program impacts.

[b] Telehealth differs from telemedicine in that it refers to a broader scope of remote health care services. Telemedicine refers specifically to remote clinical services, whereas telehealth includes these services and can include remote nonclinical services, such as provider distance-learning; meetings, supervision, and presentations between practitioners; online information and health data management; patient education; and remote admissions.

[c] A population is characterized as being socially fragile or complex or at risk for disease progression because of social circumstances or barriers. Examples include populations that are indigent, face unstable housing, and have language barriers, transportation issues, or treatment adherence problems. People with physical or mental health conditions are not considered socially fragile unless they also face social barriers, as defined above.

[d] With active enrollment (and recruitment), the awardee or its partners had direct contact with potential participants through telephone calls, mail, an arranged meeting, or a meeting triggered by a predefined event such as a hospital discharge or admission. If individuals agreed to receive services, they were enrolled into the program. With passive enrollment (and no recruitment), participants were individuals who saw a program provider and met the program eligibility requirements. Passively enrolled participants might not have been aware that they were receiving or benefiting from program services.

Appendix H:

Summary of proposed payment models and source of funding for program sustainability

This page has been left blank for double-sided copying.

Table H.1. Summary of type and status payment models and source of funding for sustaining programs, by awardee

Awardee	Proposed payment model	Was payment model implemented and used to fund program?	Source(s) of funding for sustaining program
Awardees that continued their programs largely intact (N = 11)			
AAMC	New Medicare FFS billing codes to reimburse PCPs and specialists for electronic communication between each other	No, but the awardee is in the process of negotiating with Medicare and five state Medicaid programs	Internal
Avera	PBPM fee from providers (nursing facilities) to use the eLTC service	Yes, nursing facilities pay awardee to participate in program	Payment model
CHIIC	Shared savings from rural MSSP ACO supplemented with existing Medicare FFS billing codes for population health activities, such as transitional care management (TCM) and annual wellness exams	Yes, participating sites can bill FFS but have not generated shared savings	Payment model, internal
FPHNY	One-time bundled payment for patients with HCV from Medicare and Medicaid MCOs to fund care coordination services, subject to shared savings and losses	No, but awardee began developing a new FFS payment model after progress on the originally proposed model stalled	Internal
FSCL	Multi-payer bundled payment for set of palliative care and hospice services	No, but negotiations with commercial payers in progress	Internal, existing FFS payments
Montefiore	Shared savings through the Next Generation ACO for Medicare FFS patients and value-based PBPM payment to cover care management services for others	Yes, pilot testing payment model with two health plans	Payment model, external (state DSRIP revenues)
NHCHC	Each of the 5 participating sites developed their own payment models, including PPS for FQHCs; Medicaid FFS; and flat PBPM payments from MCOs or ACOs	Yes, all 5 sites implemented multi-payer payment models	Payment model, external (grants, foundations, partnering hospitals), internal
Northwell	Condition-specific, population-based payment with value-based incentives and penalties for nephrologists	No, but continued collaborating with the National Kidney Foundation to develop payment model	Internal
UIC	PBPM fee to be paid by Medicaid MCOs for care coordination services	No, negotiations with managed care plans stalled due to delays in implementing managed care contracts by the state Medicaid agency	Internal
UKS	Medicare FFS payments from existing TCM and CCM codes; shared savings from MSSP ACO	Yes, participating sites billing Medicare FFS and ACO generated shared savings; awardee also pursuing FFS payments from commercial payers	Payment model, external (grant), internal
UNM	Medicare FFS payments from existing codes that cover neurological and neurosurgery telehealth services	Yes, and continuing negotiations with Medicaid and commercial payers for same arrangement	Payment model, internal

Table H.1 (*continued*)

Awardee	Proposed payment model	Was payment model implemented and used to fund program?	Source(s) of funding for sustaining program
Awardees that sustained parts of their program (N = 18)			
Altarum	Enhanced FFS payments from Medicaid and commercial payers for providing preventive services and dental care, and incentive payments for meeting preventive service targets	No, unable to start negotiations due to lack of data showing cost savings	External (program partner), internal
Amerigroup	None (because partner organization secured state funding early in award period)	No, did not develop payment model	External (state funding)
CCC	New Medicare FFS billing code for each patient who receives enhanced discharge planning services	No, decided not to pursue payment model after analyses showed program did not achieve intended goals	Internal
CCNC	Risk- and value-based payments combined with PBPM payments for care management and coordination services	No, but continued pursuing funding from payers	Internal (awardee consults with pharmacies, but does not provide payments to them anymore)
CHS	Medicaid MCOs pay awardee PBPM payments in exchange for primary and preventive care services	No, did not move forward with payment model and secured other funding instead	External (program partner)
Clifford Beers	Value-based arrangement with the state Medicaid agency and provider organizations to provide care coordination services; and FFS payments from commercial payers	Yes, established contract with one commercial payer, but was unable to reach agreements with Medicaid	Payment model, external (state funding)
Columbia	PBPM payments from Medicaid to dentists to cover preventive dental services	No, but continued pursuing Medicaid and other payers	External (grant)
DMC	None (did not develop payment model, but relied on existing FFS codes)	No, awardee did not develop a payment model	Existing FFS codes
Icahn	Bundled payments for acute care services that could be tailored to different commercial payers; and bundled payment with risk sharing for Medicare	Yes, executed contracts with two commercial payers, but lacked funding for patients insured by other payers (including Medicare)	Payment model
NACHRI	Per capita care management payments, shared savings, or Medicaid FFS payments (each participating site developed its own payment model)	Yes, for 5 of the 10 participating sites executed contracts with a state Medicaid agency or Medicaid MCO, while the other 5 sites used a combination of internal and external funding	Payment model, external, internal
NMC	Bundled payment from commercial payers per patient for one episode of services; FFS payments from commercial payers, Medicare, and Medicaid	No, but discussions with commercial payers in progress	Internal

Table H.1 (*continued*)

Awardee	Proposed payment model	Was payment model implemented and used to fund program?	Source(s) of funding for sustaining program
NYCH+H	Global risk-capitated contracts with two Medicaid and Medicare MCOs; value-based payment model in which Medicaid and commercial FFS payments would be adjusted for quality performance; and shared savings model with an all-payer ACO	No, did not pursue payment models after DSRIP funding became available	External (DSRIP funding), internal
UCSF	New Medicare FFS billing codes for chronic care management and advanced care planning; and a value-based payment model	No, payment model development and implementation on hold until more data on program outcomes and costs are available	External (grant)
UHCMC	CMS' Oncology Care Management (OCM) payment approach, which combines FFS payments with enhanced PBPM payments from Medicare and participating commercial payers; and a coordination fee and shared savings arrangement through an ACO	Yes, some funding through OCM, but negotiations with ACO delayed due to ACO's leadership changes	Internal, external (program partner)
UMich	Provider incentives to surgeons for meeting process measure benchmarks; and FFS billing with commercial payers	Yes, implemented FFS billing codes with one commercial payer	New FFS billing with commercial payer, external (program partner)
WashU	Bundled payment for 90-day episode of contraceptive care by Medicaid and commercial payers	No, negotiations with the two payers stalled due to changes in payer leadership	Existing billing mechanisms (but will have to discontinue services that cannot be billed)
WI DHS	One-time payment for enrolling patient; ongoing capitated payments for care coordination	Yes, the Medicaid state plan amendment made billing codes available for both types of payments	Payment model
Yale	Prospective, population-based payments to EMS medical directors by commercial and Medicare payers for beneficiaries in a region at risk for falling	No, unable to engage payers	Internal (continuing modified version of program)
Awardees that ended their programs after the end of the award (N = 9)			
ACCF	Bundled payments from commercial payers to support diagnosing and treating patients with stable ischemic heart disease	No, unable to advance payment model due to challenges to implementing program and accessing data needed to calculate payment amounts	None
BMC	PBPM payments from Medicaid ACOs that cover care coordination services	No, negotiations stalled so awardee ended program after award funding ended	None
Hopkins	Risk-adjusted PBPM payments for memory care coordination paid by organizations that bridge Medicare and Medicaid (e.g. PACE programs, Medicaid ACOs, and Integrated Care Organizations)	No, but continued developing business case for payment model	None
Mesa	Existing and new FFS billing codes for all payers	No, unable to engage payers due to insufficient incentives and lack of data to prove cost savings	None

Table H.1 (*continued*)

Awardee	Proposed payment model	Was payment model implemented and used to fund program?	Source(s) of funding for sustaining program
SCH	Medicaid MCOs pay a PBPM care management fee adjusted for measures of quality, use, and spending	No, unable to reach agreements with Medicaid MCOs	None
UCSD	PBPM payments that cover care management services paid by Medicaid Advantage, commercial, and Medicaid plans	No, but continued negotiations with payers	None
UNC	Episode-based bundled payments for Medicare FFS, Medicare Advantage, and Medicaid patients with acute, nonspecific lower back pain	No, discontinued developing payment model due to challenges accessing data needed to calculate payment amounts	None
VCC	PBPM payment to cover health coaching services for patients with HIV, paid by ACOs or Medicaid MCOs	No, but continued negotiations with Medicaid MCOs	None
Ventura	Discounted bundled payments for treating COPD patients, and incentive payment for meeting process measures treated to treating COPD, paid by Medicare	No, unable to engage Medicare	None

Mathematica
Progress Together

Evaluation of the Health Care Innovation Awards, Round 2: Final Report

Part II: Individual Awardee Reports

List of HCIA R2 Awardees (in alphabetical order)
Altarum Institute
American College of Cardiology Foundation
Amerigroup
Association of American Medical Colleges
Avera Health
The Board of Trustees at The University of Illinois, Chicago
Boston Medical Center
CareChoice Cooperative
Catholic Health Initiatives Iowa Corporation
Children's Home Society of Florida
City of Mesa Fire and Medical Department
Clifford W. Beers Guidance Clinic
Community Care of North Carolina
Detroit Medical Center
Four Seasons Compassion for Life
The Fund for Public Health in New York
Icahn School of Medicine at Mount Sinai
Johns Hopkins University
Montefiore Medical Center
National Association of Children's Hospitals and Related Institutions
National Health Care for the Homeless Council
Nebraska Medicine
New York City Health + Hospitals
Northwell Health
Regents of the University of California at San Diego
Regents of the University of California at San Francisco
Regents of the University of Michigan
Seattle Children's Hospital
Trustees of Columbia University in the City of New York
University Hospitals Cleveland Medical Center
University of Kansas Health System
University of New Mexico
University of North Carolina at Chapel Hill
Ventura County Health Care Agency
Village Center for Care
Washington University School of Medicine
Wisconsin Department of Health Services
Yale University

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Altarum Institute

September 2020

Kristin Andrews Lemos, Dana Petersen, Theresa Feeley-Summerl, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

ALTARUM INSTITUTE

Altarum Institute (Altarum), a nonprofit health research organization, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Michigan Caries Prevention Program (MCPP). Altarum created the program in partnership with Delta Dental of Michigan, the University of Michigan (UM) School of Dentistry, and the Michigan Department of Health and Human Services (DHHS). The MCPP sought to address a critical care gap in preventive early childhood dental care and to encourage establishing dental homes earlier in childhood than has historically occurred for Medicaid and Children's Health Insurance Program (CHIP) beneficiaries in the state. thereby reducing the incidence of dental disease and related costs. The MCPP launched in May 2015, nine months after award. The intervention period funded by HCIA R2 ended in February 2018. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The MCPP, a provider-level intervention without direct provision of services to beneficiaries, aimed to improve dental and oral health outcomes for children and reduce associated Medicaid costs through three components: (1) training and TA to PCPs to deliver preventive oral health care in the primary care setting and refer children to dentists, (2) a new health IT system to help coordinate referrals between PCPs and dental providers, and (3) participant and family engagement.
- This evaluation used a longitudinal difference-in-differences model to estimate the impact of the training and TA component on the receipt of dental and oral health care and dental and oral health outcomes. Due to data limitations, the impact of the intervention on costs could not be estimated. According to the awardee's theory of action, oral health interventions are unlikely to affect the Centers for Medicare & Medicaid Services' core outcomes of ED visits and hospital admissions, and thus total cost of care.
- The impact analysis examined the training and TA component of the intervention only, and includes 94,944 Medicaid beneficiaries attributed to 812 PCPs who participated in the training and TA component of the program, and 124,696 beneficiaries attributed to 2,281 comparison PCPs who did not participate in the training and TA.

Altarum and its partners hypothesized that by improving the ability of primary care providers (PCPs), dental providers, and public health providers to deliver evidence-based oral health care to young children, children would receive preventive oral health care and education at a younger age, ultimately reducing the amount of dental disease, the number of dental caries, and associated costs. The three intervention components were (1) training and technical assistance (TA) provided to PCPs (which included pediatricians, family practice specialists, and other internal medicine specialists) and their office staff to build their capacity to deliver evidence-based preventive oral health services in the primary care setting and refer patients to dental homes, (2) a health information technology (health IT) system to facilitate communication and help coordinate referrals between PCPs and dental providers, and (3)

participant and family engagement. To achieve its goal of improving the receipt of preventive dental and oral health services in early childhood, the program aimed to (1) expand delivery of preventive oral health services by PCPs who serve children, (2) integrate oral health care across primary care and dental settings, and (3) build the capacity of the oral health safety net. The MCPP did not provide services directly to patients; rather, the program trained PCPs, dental providers, staff at Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) clinics, and others who serve the target population. The awardee considered all children served by participating providers as indirect participants.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The MCPP sought to improve access to preventive oral health and dental services and encourage establishing dental homes for the youngest Medicaid and CHIP beneficiaries in Michigan, thereby reducing the amount of dental disease, the number of dental caries, and associated costs.
Major innovation	The MCPP was innovative because it sought to improve access to and delivery of preventive oral health care through a variety of strategies, including increasing the capacity of PCPs to provide preventive oral health services in the primary care setting; using health IT to improve information sharing and referrals between medical and dental providers; and increasing the capacity of public health professionals, educators, and caregivers to promote children's oral health.
Program components	• Training and TA to PCPs serving children providing evidence-based education on directly providing oral health services in the primary care setting • A health IT system designed to allow information sharing between PCPs and dental providers; participant and family engagement, including educational outreach to dentists, dental hygienists, and public health professionals; training for staff at WIC clinics; use of broad-based dissemination strategies such as conference presentations and social media; and a crowdsourcing website (SmileConnect) to help fill needs for early dental health services and supplies
Target population	Medicaid and CHIP beneficiaries in Michigan up to the age of 17 years[a]
Total enrollment of providers	MCPP was a provider-level intervention without direct provision of services; all patients served by a participating provider were considered indirect participants. 1,565 PCPs (104 percent of the awardee's enrollment target for providers) participated in the training and TA component of the intervention. The awardee also trained 1,588 nonprovider support staff in the participating practices. An estimated 9 percent of eligible PCPs in Michigan participated in the training and TA component.[b]
Total enrollment of beneficiaries	An estimated 949,164 Medicaid and CHIP beneficiaries (95 percent of awardee's original enrollment target of 1 million) indirectly participated in the intervention across all three intervention components. An estimated 157,985 beneficiaries (an estimated 12 percent of all Medicaid and CHIP-enrolled children who visited a PCP in Michigan during the follow-up period) indirectly participated in the intervention's training and TA component (that is, ever visited a trained provider during the follow-up period).
Level of engagement	All beneficiaries included in the impact analysis saw their PCPs at least once during the baseline and follow-up periods, respectively, and therefore had the opportunity to receive the preventive oral health services and education targeted by the intervention's training and TA component and to receive a referral to a dental home.

Table 1 (*continued*)

Program characteristics	Description
Theory of change or theory of action	Altarum and its partners hypothesized that by improving the ability of PCPs and other providers to deliver evidence-based oral health care to young children, children would receive preventive dental and oral health care (including fluoride varnish, oral health evaluation, dental sealants, and preventive dental visits) at a younger age, ultimately reducing the incidence of dental disease (requiring restorative procedures or in some cases, dental-related ED visits) and associated costs.
Award amount	$9,383,762
Effective launch date	May 8, 2015
Program settings	The intervention's training and TA component targeted pediatric-focused primary care clinics and medical centers; other intervention components targeted dental offices and other community, early education, school, and public health settings.
Market area	Urban, suburban, and rural Michigan (all regions of the state)
Target outcomes	• Increase the proportion of children in low-income households who receive preventive oral health care by 60 percent • Reduce the proportion of Medicaid and CHIP beneficiaries who have dental caries by 30 percent • Provide a net savings to CMS of $21.1 million
Payment model	Altarum proposed a FFS payment model that would cover the cost of the program for a subset of its program population. For children younger than 3, PCPs and dentists would be eligible for enhanced FFS payments for providing preventive services and dental care, respectively. Both types of providers could earn bonus payments if they met targets for increasing preventive services for children. However, the awardee discontinued plans to engage payers after finding insufficient evidence of cost savings.
Sustainability plans	The awardee discontinued training and TA to PCPs in May 2017. Altarum sustained the SmileConnect site with its own funding, while pursuing longer-term funding for the site from state- and locally based payers and foundations. A national oral health organization took over the provision of supplies. The awardee replicated the MCPP in Los Angeles with funding from California's Section 1115 waiver demonstration. The awardee also obtained funding to expand the WIC training into rural areas and sought funding to expand use of the health IT program component into the child welfare system.

[a] Altarum lowered its enrollment goal from slightly more than 1 million indirect participants to 742,715 indirect participants in the third program year. This count includes all indirect participants (beneficiaries) served by the intervention's three components, not just those affected by the training and TA component examined for the impact evaluation. When providers signed up for MCPP's training and TA component, they estimated the practice's caseload of Medicaid and CHIP-enrolled children ages 17 and younger. The total number of indirect participants served includes the estimated practice caseload, plus participants indirectly served by WIC providers, as well as those served through SmileConnect, the awardee's crowdsourcing website.

[b] According to CMS NPPES and Michigan Medicaid claims data, there were about 10,027 PCPs in Michigan who saw any Medicaid or CHIP-enrolled children in the intervention's baseline year time frame; 901 (9%) participated in the training and TA component of the intervention.

CHIP = Children's Health Insurance Program; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; IT = information technology; MCPP = Michigan Caries Prevention Program; NPPES = National Plan and Provider Enumeration System; PCP = primary care physician; TA = technical assistance; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

The impact analysis presented in this report examines the training and TA component of the intervention only and included only the 94,499 Medicaid and CHIP beneficiaries attributed to PCPs who participated in the training and TA component of the program. The comparison group included 124,696 Medicaid and CHIP beneficiaries attributed to a matched comparison group of eligible PCPs located in similar geographic areas in Michigan who did not participate in the training and TA. Because MCPP was a provider-level intervention without direct provision of services to beneficiaries, a list of participating beneficiaries was not available from the awardee. Instead, beneficiaries were attributed to PCPs for the impact analysis. Of all eligible PCPs in Michigan, the estimated participation rate of providers in the intervention's training and TA component was 9 percent; the intervention's training and TA component indirectly affected all of the beneficiaries who visited one of these PCPs. (These PCPs served about 12 percent of eligible Medicaid and CHIP beneficiaries – those with a visit to a PCP- in the state of Michigan.) Table 2 summarizes the key features of the impact evaluation. Appendix A, Tables A.1 and A.2 describe the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a longitudinal difference-in-differences model to estimate the impact of the intervention's training and TA component on the receipt of dental and oral health care and dental and oral health outcomes. Because MCPP was a provider-level intervention without direct provision of services to beneficiaries, a list of participating beneficiaries was not available from the awardee. Instead, the impact analysis includes beneficiaries attributed to treatment or comparison PCPs based on where they received the plurality of their primary care visits.[b]
Intervention group for evaluation	The impact analysis examines only the training and TA component of the intervention and includes 94,944 Medicaid beneficiaries attributed to 812[a] PCPs who participated in the training and TA component of the program.
Comparison group	The impact analysis compared outcomes among participants to those of a comparison group of 124,696 beneficiaries attributed to 2,281 comparison PCPs who did not participate in the training and TA.
Limitations	If participants differed from eligible nonparticipants in ways not captured in Medicaid claims data, the impact estimates might be biased. The low participation rate (about 9 percent of eligible PCPs) would have made it difficult to identify impacts if measured over all eligible beneficiaries. Conversely, measurable program impacts could have been larger if measured only among the subset of children (approximately half of the treatment group) that did not have an already-established dental home.

[a] Excludes 753 PCPs who did not see any Medicaid or CHIP beneficiaries during the baseline year or who were dropped during the matching, attribution, or modeling steps. See Appendix A for full details.

[b] Beneficiaries included in the impact evaluation could be attributed to a different provider at baseline versus follow-up; however, they must have been attributed to a treatment or comparison provider at both baseline and at follow-up for inclusion in the analysis. Beneficiaries included in the comparison group were required to have no visits with any treatment provider during the follow-up period to ensure that the training and TA component of the intervention did not affect them.

MCPP = Michigan Caries Prevention Program; PCP = primary care physician; TA = technical assistance.

PROGRAM DESIGN AND ADAPTATION

Through MCPP, Altarum staff and its partners delivered three key intervention components: (1) training and TA to PCPs and their office staff, (2) implementation of a new health IT system to facilitate information sharing and referrals between PCPs and dentists, and (3) participant and family engagement.[1]

Training and TA

Beginning in the first program year, the awardee offered in-person training to PCPs serving children and their office staff that covered evidence-based standards of oral health care. This one-time training sought to enable PCPs to directly provide oral health services in the primary care setting and covered (1) evidence-based standards of preventive oral health care, (2) oral health screening, (3) referrals to dentists and establishing a dental home, (4) applying fluoride varnish, (5) patient and family education on oral health care, (6) processes for obtaining Medicaid reimbursement for covered oral health services, (7) guidance on how to adapt the intervention to the provider site, and (8) training on the use of the health IT system (described below). This training aimed to equip PCPs to adopt and hone their skills in evidence-based oral health practices (oral health screening, applying fluoride varnish, and oral health education); integrate these practices into their usual care delivery; and refer children to dentists and coordinate with them to establish dental homes for children. Altarum provided ongoing TA for four to seven months after the initial training.

Implications of program implementation for detecting impacts

- The awardee engaged PCPs serving children in nearly all regions of the state and delivered a high-quality training and TA program focused on the providing oral health services in the primary care setting. As a result, the intervention had potential to improve children's receipt of preventive oral health services.
- The awardee encountered delays in the implementation of the health IT system (MiDR) which likely hampered information sharing and referrals between PCPs and dentists. In addition, particularly in the intervention's first year, providers encountered problems billing managed care organizations for these services, which might have discouraged some providers from continuing to offer these services.
- Dental providers' limited engagement with MCPP and low rates of MiDR use diminished the program's effectiveness in establishing dental homes for children and in reducing the incidence of dental caries and dental disease.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the MCPP program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Health IT

The awardee developed a public health improvement and health IT system for medical and dental providers to document oral health services provided to children and refer patients to dentists and dental specialists accepting Medicaid. Known as the Michigan Dental Registry (MiDR), the health IT system included a web-based interface and an electronic medical record (EMR) interface to facilitate documentation and referral. The EMR module also provided decision support and included a screening tool for oral health risks. The awardee offered training to both medical and dental providers on the use of MiDR.

Engaging participants and families

The participant and family engagement component included educational outreach to dentists, dental hygienists, and public health professionals; training for staff at WIC clinics; use of broad-based dissemination strategies such as conference presentations and social media; and a crowdsourcing website (SmileConnect) to help fill needs for early dental health services and supplies. As part of these engagement strategies, Altarum and its partners educated dental providers, via conferences and presentations, about the importance of oral health care for young children and provided tips on delivering dental care and oral health education to children and their families.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee successfully engaged providers in nearly all regions of the state and delivered a high-quality training and TA program to their office staff and the participating PCPs serving children. As a result, the intervention had the potential to increase children's access to preventive oral health services. The training and TA program aimed to train and support PCPs in providing oral health services in the primary care setting and to receive Medicaid reimbursement for doing so. The program curriculum, based on Smiles for Life (an online training program for PCPs to become certified to provide—and in Michigan, to bill for—oral health services), trained PCPs on evidence-based standards of preventive oral health care. The MCPP built upon this curriculum with the added value of in-person training on administering fluoride varnish, administrative training on the billing process, a workflow assessment to help clinics fit oral health screening and fluoride varnish application into a short patient appointment, providing free starter supplies, and MiDR training when it was implemented.

To encourage provider participation and engagement, the program offered continuing medical education (CME) and maintenance of certification (MOC) Part IV credits to PCPs participating in the training and TA component. Participants also came out of the training certified to receive Medicaid reimbursement for providing oral health services (including fluoride varnish and oral health evaluation) in the primary care setting, and office staff came away trained on how to bill for these services. Finally, the awardee leveraged oral health champions—one or two motivated

individuals within each participating primary care practice tasked with promoting the training to colleagues and sustaining the program within the practice following the training and TA—to encourage providers' participation and improve engagement.

In spite of these achievements, several implementation challenges might have made it more difficult for the intervention to achieve some of its intended goals. Particularly in the intervention's first year, PCPs encountered challenges receiving reimbursement from Medicaid managed care organizations (MCOs) for oral health services. By summer 2016, program staff were confident that this issue was mostly resolved. Furthermore, implementing one of the intervention's three components- the MiDR health IT system- was delayed until fall 2016. This meant that training for the first set of clinics did not include MiDR-specific training, and those clinics could not use the health IT system in the intervention's first year or more, which likely limited referrals and coordination with dental providers during this time frame. After the awardee implemented MiDR in October 2016, the awardee circled back to provide MiDR-specific trainings to about half of previously trained sites. Although the awardee successfully engaged PCPs through the training and TA component and aimed to engage dental providers through several means, the program achieved limited engagement with dental providers. The historical lack of linkages and collaboration between dental providers and PCPs, and the delayed implementation of MiDR (which sought to help overcome this issue), posed a barrier to the awardee's efforts to improve information sharing and referrals between PCPs and dentists.

ESTIMATING PROGRAM IMPACTS

Study sample

Because MCPP was a provider-level intervention without direct provision of services to beneficiaries—and therefore, the awardee did not have data on beneficiary indirect participants—beneficiaries were attributed to providers for the impact analysis. The study is based on 94,944 Medicaid and CHIP beneficiaries attributed to 812 treatment PCPs (excluding the 753 participating PCPs who did not have any visits with Medicaid beneficiaries during the study period), and 124,696 Medicaid and CHIP beneficiaries attributed to 2,281 matched comparison providers. The analysis matched treatment providers to similar comparison providers in the same county or region, but in different zip codes, which ensured that county-level factors were similar for both groups; at the same time, restricting matches to outside the zip code prevented matching a comparison provider to a treatment provider from the same practice.

The awardee delivered the training and TA component of the intervention only to participating PCPs and their office staff; that is, those who served the attributed children who formed the intervention's treatment group. The other two intervention components (health IT and patient and family engagement) had the potential to affect low-income children in Michigan more broadly. As a result, the impact analysis in this report could evaluate the impact of only the training and TA component to PCPs on children's receipt of preventive care and health outcomes.

Characteristics of treatment and comparison group beneficiaries

Characteristics of the Medicaid and CHIP beneficiaries attributed to treatment and comparison providers were fairly well balanced at baseline (Table 3). (Appendix B provides balance results from provider-level matching, measured during the 12 months before enrollment.) Average age during the baseline year was 6.1 years for treatment group children and 6.5 years for comparison group children, which reflects the intervention's focus on children ages birth to 17 and particularly on the youngest subset (up to age 3). The beneficiaries in both groups largely reflected the demographics of the overall Medicaid-enrolled population in Michigan (Kaiser Family Foundation 2019). A slightly higher percentage of treatment beneficiaries from urban areas compared with those in the comparison group reflects the fact that the awardee sought to engage PCPs in several urban areas for the training and TA component.

Baseline health status and service use were very similar across groups. The mean baseline Chronic Illness and Disability Payment System (CDPS) score for treatment and comparison group children was very low at 1.0, reflecting that this population was, on average, relatively free from chronic conditions or disability. More than one-fifth of each group had a restorative procedure (such as a filling or a root canal) in the baseline year. Slightly more than half of each group had a preventive dental visit and about half of both groups received fluoride varnish during the baseline year, indicating that these beneficiaries might have had an already-established dental home at baseline.

Table 3. Baseline characteristics of MCPP treatment and comparison group beneficiaries

Measure	Treatment (N = 94,944)	Comparison (N = 124,696)
Demographics		
Age at enrollment, years	6.1	6.5
Age group, %		
0 to 3 years	36	33
4 to 6 years	23	22
7 to 12 years	31	33
13 to 17 years	11	13
Male, %	51	51
White, %	59	57
Black, %	22	22
Hispanic, %	8.5	9.5
American Indian, Alaska Native, Asian/Pacific Island American, or other, %	0.75	0.89
Unknown, %	9.8	11
Residence		
Urban	77	84
Suburban	20	13
Rural	2.4	3.1

Table 3 (*continued*)

Measure	Treatment (N = 94,944)	Comparison (N = 124,696)
Health status and acute and preventive service use during the year before enrollment		
CDPS score[a]	1.0	1.0
Any ambulatory sensitive ED visit for dental caries, %	0.23	0.21
Any restorative procedure, %	22	22
Any preventive dental visit, %	54	53
Any receipt of fluoride varnish, %	51	50

Source: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

Notes: The baseline period covers a 12-month provider-specific period.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid.

Appendix B presents balance results from provider-level matching.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; MCPP = Michigan Caries Prevention Program.

Analytic approach

The analysis used a longitudinal difference-in-differences study design to estimate impacts. This design measures program effects as the change in outcomes among beneficiaries attributed to study PCPs before versus after enrollment relative to the change in outcomes among beneficiaries attributed to a matched comparison group of PCPs (who had similar characteristics to treatment PCPs) over the same period. The analysis also used beneficiary-level fixed effects to control for unobservable beneficiary-level characteristics that were constant over time. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects.

The primary outcomes include the number of preventive dental visits, oral health evaluations, and fluoride varnish applications. Secondary outcomes include the number of ED visits, ambulatory-sensitive ED visits for dental caries, and restorative procedures. The number of hospital admissions could not be examined because the Medicaid claims and encounter data files received for the evaluation did not include a complete set of hospital claims. Furthermore, cost data were not available because most children in the study population were enrolled in Medicaid managed care.

The study defined the pre-enrollment period as the year before each participating PCP's enrollment date, with the enrollment date defined as the date on which the PCP received the training and TA component of the intervention. It defined the post-enrollment period as the two years following the PCP's enrollment date. A pseudo-enrollment date was assigned to each comparison PCP within one month of the enrollment date of the treatment PCP to which he or

she was matched. Appendix A describes the data, statistical models and outcomes used to estimate the effects of the program, as well as the identification of the final analytic sample.

IMPACT RESULTS

The training and TA component of the MCPP had a discernible favorable impact on three of the key outcomes that the awardee expected to influence according to its theory of action: a **15 percent** estimated increase in receipt of fluoride varnish, a **2.6 percent** estimated increase in receipt of oral health evaluations, and a **2.8 percent** estimated increase in the rate of preventive dental visits when measured over the cumulative two-year follow-up period (Table 4). For these three measures, the MCPP also had a discernable, favorable impact individually in each of the two follow-up years. These findings are robust; results were similar for a more inclusive, cross-sectional model that included beneficiaries excluded from the primary model (those who visited a treatment or comparison PCP at either baseline or follow-up, but not both) (Appendix C, Table C.3 presents results). Conversely, there was no discernable impact on receipt of dental sealants or on rates of restorative procedures, services typically administered only by dental providers. According to its theory of action, the awardee expected to affect these outcomes by PCPs referring more children to dental providers, and (for restorative procedures specifically) by PCPs providing preventive oral health services through the intervention's training and TA component.

Finally, the intervention did not show a discernable impact on the subset of ED visits that the intervention had the greatest likelihood of affecting, according to its theory of action—ambulatory-sensitive ED visits for dental caries. Appendix C presents full results, including estimated impacts on all ED visits.

Table 4. Estimated percentage impact of MCPP on selected outcome measures

	Year 1	Year 2	Cumulative
Fluoride varnish, per 1,000 beneficiaries			
Impact (count)	37***	27***	29***
Percentage impact	19%	8.4%	15%
p-value	< 0.01	< 0.01	< 0.01
Oral health evaluations, per 1,000 beneficiaries			
Impact (count)	3.6***	7***	5.1***
Percentage impact	1.9%	3.5%	2.6%
p-value	< 0.01	< 0.01	< 0.01
Preventive dental visits, per 1,000 beneficiaries			
Impact (count)	10***	17***	13***
Percentage impact	2.2%	3.6%	2.8%
p-value	< 0.01	< 0.01	< 0.01
Restorative procedures, per 1,000 beneficiaries			
Impact (count)	-2.3	-2.0	-2.0
Percentage impact	-1.5%	-1.2%	-1.3%
p-value	0.42	0.54	0.43

Table 4 (*continued*)

	Year 1	Year 2	Cumulative
Dental sealants, per 1,000 beneficiaries			
Impact (count)	1.9	-0.22	0.89
Percentage impact	5.1%	< 1%	2.6%
p-value	0.12	0.86	0.41
Sample sizes			
Treatment	87,628	82,004	94,944
Comparison	116,071	99,456	124,696

Sources: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

Notes: Impact estimates for the number of visits are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate).The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment. Appendix C presents full impact estimates.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; MCPP = Michigan Caries Prevention Program.

The high quality of the training and TA program facilitated MCPP's ability to demonstrate discernable, favorable impacts on the receipt of fluoride varnish, oral health evaluations, and preventive dental visits. Altarum staff described how the program fully supported PCPs in becoming certified to receive Medicaid reimbursement for oral health services, offered administrative training on the billing process, provided a workflow assessment to help clinics fit oral health screening and fluoride varnish applications into a brief patient appointment, and gave practices free starter supplies. These supports and supplies helped PCPs to immediately implement the provision of preventive oral health services into their daily practice and to receive Medicaid reimbursement for doing so. Altarum leaders noted that the quality improvement requirements of the CME and MOC Part IV program, which required PCPs participating in this component to submit monitoring data at baseline and again at four and seven months after the training date, supported their ability to provide high quality TA tailored to the specific needs of each participating practice. Altarum staff reviewed monitoring data that the practices submitted four months after the training date to identify areas for targeted TA to each practice and analyzed data submitted seven months after the training date to identify facilitators of and barriers to practices' success and refine the approach for future trainings. Nearly all clinician respondents (93 percent) surveyed in the second half of the third program year strongly or somewhat agreed that they learned new skills that were important for their roles, and 89 percent strongly or somewhat agreed that the training helped them improve their job performance. In addition, most (96 percent) indicated the program was very or somewhat effective in achieving its goals.

In addition, the MCPP's use of a multifaceted strategy to recruit and engage PCPs serving children, including a strong value proposition for participating in the intervention, facilitated providers' participation and engagement. Even though only an estimated 9 percent of eligible PCPs in the state participated in MCPP, there was a large study sample of Medicaid- and CHIP-

enrolled children attributed to participating PCPs; this meant there was ample statistical power to detect program impacts, even ones that were relatively small.

Main findings from impact evaluation

- The MCPP's training and TA component increased receipt of fluoride varnish, receipt of oral health evaluations, and preventive dental visits during the two-year follow-up period.
- There was no discernable increase in receipt of dental sealants or reduction in rates of restorative procedures or ambulatory-sensitive ED visits for dental caries during the two-year follow-up period.

However, the fact that the three estimated program impacts (receipt of fluoride varnish, receipt of oral health evaluations, and preventive dental visits) were relatively small to modest might be related to several challenges the awardee and participating providers faced. Particularly in the program's first year, PCPs experienced challenges obtaining reimbursement from Medicaid MCOs for oral health services. Although the awardee and Michigan DHHS worked with Medicaid MCOs to align with state Medicaid reimbursement policy, these challenges might have dissuaded some providers (particularly those trained early in the program) from staying engaged and continuing to provide these services throughout the award period. Furthermore, despite resolving the reimbursement issues, the awardee noted that Michigan Medicaid reimbursement for fluoride varnish and oral health evaluation was low compared to other states and commercial plans, which might also have weakened some providers' incentives for staying engaged and continuing to offer these services. In addition, the delayed implementation of MiDR meant that the first set of trained sites did not have the opportunity to use the system to refer to and coordinate with dental providers in the first 15 months of the award period. The results from a survey of PCPs conducted in the third program year underscored the implications of the delayed MiDR implementation; only 30 percent of PCPs reported that they spent more time in a referring children to dentists and following up on referrals compared to a typical week before they participated in MCPP and only 10 percent said they spent more time coordinating care and communicating with dentists.

These relatively low rates of coordination with dental providers might have also contributed to the lack of impacts on two of the measures administered by dental providers—receipt of dental sealants and rates of dental treatments (restorative procedures). The historical lack of linkages and collaboration between PCPs and dental providers, and the delayed implementation of MiDR (which aimed to help overcome these factors), posed a barrier to the awardee's efforts to increase information sharing and referrals between PCPs and dentists. It is also important to note that the benefits of improved preventive dental and oral health care on the dental and oral health outcomes of pediatric populations (for example, incidence of dental disease, dental caries, and related need for restorative procedures) are typically not observed in the short term and tend to be longer-term outcomes. Therefore, it is possible that any potential impacts of the improved preventive care on the rates of restorative procedures are more likely to appear after the two-year follow-up period.

CONCLUSION

The MCPP demonstrated relatively small to modest favorable estimated impacts on receipt of preventive oral health services and preventive dental visits among Medicaid- and CHIP-enrolled children. The high quality of the MCPP's training and TA program, and the awardee's multifaceted provider engagement strategy, which enabled the training and TA component to reach a large number of providers and Medicaid- and CHIP-enrolled children, facilitated these favorable estimated impacts. However, the fact that these estimated impacts were relatively modest could in part reflect the delayed implementation of MiDR (which posed a barrier to improved information sharing and referrals between PCPs and dentists), as well as the reimbursement-related challenges PCPs encountered in the intervention's first year. Moreover, the program demonstrated no discernable impacts on two measures typically administered by dental providers: dental sealants and restorative procedures. Although the goal of the MCPP's training and TA component was to engage PCPs, these findings underscore the need to also effectively engage and incentivize dentists as active partners in improving children's dental and oral health.

Limitations of evaluation

Providers' participation in the MCPP's training and TA component was voluntary, and only 9 percent of eligible PCP's participated. Lack of information on the participations' decisions might mean that treatment providers differed from the comparison providers in unobservable ways that could affect outcomes. In addition, because data linking comparison providers to practices were unavailable, it was not possible to confirm that all comparison providers were drawn from practices that did not receive the training. To minimize the risk of drawing comparison providers from practices that received the training, matched comparison providers were drawn from the same county or region but a different practice location zip code than the treatment counterparts. Furthermore, children attributed to comparison providers were not allowed to have a visit in the follow-up period with any trained treatment provider to help ensure that children in the comparison group never had the opportunity to receive the intervention. Finally, slightly more than half of each group had a preventive dental visit and about half of both groups received fluoride varnish during the baseline year, indicating that these beneficiaries might have had an already-established dental home. Measurable program impacts could have been larger among the subset of children without an already-established dental home.

PROGRAM SUSTAINABILITY

At the end of its award in February 2018, Altarum sustained several aspects of the MCPP with a combination of external funding, its own funding, and help from its partners. Specifically, a national oral health organization took over providing supplies. Altarum funded the SmileConnect site with its own resources while pursuing longer-term funding from state- and locally based payers and foundations. The awardee also stopped providing direct training and TA to providers after the award ended, but applied to the state to allow providers to obtain CME accreditation for the training from the state medical society. In addition, Altarum scaled and replicated the program to new populations. For example, the awardee obtained funding to expand the WIC training into rural areas and sought funding to expand use of MiDR into the child welfare system. Finally, Altarum replicated the MCPP in Los Angeles with funding from California's Medicaid Section 1115 demonstration.

Altarum's proposed payment model

Altarum proposed a FFS payment model that would cover the cost of the program for a subset of its program population. For children younger than 3, PCPs and dentists would be eligible for enhanced FFS payments for providing preventive services and dental care, respectively. Both types of providers could earn bonus payments for meeting targets to increase preventive services for children.

Altarum proposed a fee-for-service FFS payment model that would cover the cost of the program for a subset of its program population. For children younger than 3, PCPs and dentists would be eligible for enhanced fee-for-service payments for providing preventive services and restorative procedures, respectively. Both types of providers could earn bonus payments for meeting targets to increase preventive services for children. However, Altarum made little progress advancing its proposed payment model. The awardee did not determine which services would be eligible for enhanced payments, nor how to set the targets for receiving bonus payments. The awardee said that given the available data, it has been unable to show that the program generated Medicaid savings, which hampered its ability to justify increasing rates and bonuses.

REFERENCE

Kaiser Family Foundation. "Medicaid Enrollment by Race/Ethnicity, FY 2013." Washington, DC: Kaiser Family Foundation, 2019. Available online.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The analysis based the primary impact estimates for expenditures and number of visits on a difference-in-differences approach with beneficiary fixed effects. The estimates show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries), which are in turn defined by the enrollment date (or pseudo-enrollment date) of the primary care provider to whom he or she was attributed. Descriptions of data and outcome measures follow. Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy.

Only 52 percent of participating primary care providers (PCPs) and an estimated 60 percent of total beneficiaries indirectly participating in the intervention were included in the impact analysis (Tables A.1 and A.2). Total beneficiary participants were defined as all Medicaid and Children's Health Insurance Program (CHIP) beneficiaries having at least one in-person visit with a participating PCP during the follow-up period (May 2015 through May 2017). The study dropped most of the excluded PCPs because they did not see any Medicaid- or CHIP-enrolled children in the baseline year.

Table A.1. Identification of final sample for impact analysis for MCPP for PCPs

	Number of PCPs removed from analytic sample	Number of PCPs remaining in analytic sample
Total program participants (training and TA component) through August 31, 2017		**1,565**
Estimated number of PCP participants who could not be matched[a]	669	896
Estimated number of PCP participants with no beneficiaries attributed to them[b]	26	870
PCPs dropped due to a Flint or Genesee county business practice location	39	831
Number dropped from analysis due to switching from treatment to comparison status (or vice versa) during follow-up period	1	830
Number dropped from fixed effects model (due to only having beneficiaries attributed to them at baseline or follow-up)	18	812
Final analytic sample		**812**

Sources: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

[a] PCP did not see any Medicaid-enrolled children in the baseline year (664 PCPs) or lacked exact matching variables (5 PCPs).

[b] For instance, because the PCP did not provide the plurality of primary care to any Medicaid-enrolled children during the year.

MCPP = Michigan Caries Prevention Program; PCP = primary care provider; TA = technical assistance.

Table A.2. Identification of final sample for impact analysis for MCPP for beneficiaries

	Number of beneficiaries removed from analytic sample	Number of beneficiaries remaining in analytic sample
Total program participants (training and TA component) through August 31, 2017		**157,985[a]**
Beneficiaries not attributed to a participating PCP at either baseline or follow-up[b]	4,222	153,763
Attributed beneficiaries dropped due to their PCP having a Flint or Genesee county business practice location	13,746	147,017
Number dropped from analysis due to missing key analytic information (such as date of birth)	46	139,971
Number dropped from analysis due to switching from treatment to comparison status (or vice versa) during follow-up period	1,519	138,452
Number dropped because all beneficiary-quarters had zero eligible days	1,315	137,137
Number dropped from fixed effects model (due to not being attributed to a PCP at both baseline and follow-up)	42,193	94,944
Final analytic sample		**94,944**

Sources: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

[a] An estimated 949,164 beneficiaries indirectly participated in MCPP across all three intervention components. Altarum did not provide an estimated count of the subset of beneficiaries participating specifically in the intervention's training and TA component. Based on Michigan Medicaid data, an estimated 157,985 Medicaid and CHIP beneficiaries indirectly participated in the intervention's training and TA component by visiting a treatment PCP during the follow-up period.

[b] For instance, because the beneficiary did not receive the plurality of his or her primary care from a participating provider.

MCPP = Michigan Caries Prevention Program; PCP = primary care provider; TA = technical assistance.

Data and outcomes

The study constructed the provider-level matching file, attribution file, and beneficiary-level analysis file using Medicaid claims, encounter, and eligibility data provided by Altarum (from the Michigan Department of Health and Human Services), the awardee's provider-level finder file, and publicly available National Plan and Provider Enumeration System (NPPES) data.

The analysis constructed provider-level matching variables based on provider-level characteristics (for instance, provider's gender, credentials, taxonomy, and business location address) contained in NPPES data and Medicaid claims and encounter data on service use of each provider's patient panel. The analysis matched treatment providers to comparison providers within the same county or region, but different zip code. This method ensured that county-level factors (such as the roll-out of the Healthy Kids Dental [HKD] program, Michigan's Medicaid managed care program for children) were similar for the treatment and comparison group; at the same time, restricting matches to outside the zip code prevented matching a comparison provider to a treatment provider from the same practice. Some counties—primarily those in which the HKD rollout did not coincide with the intervention period or the baseline year—were aggregated

into small regions based on geography and median household income. In addition, because the Flint water crisis evolved concurrently with the intervention (and exposure to lead has been shown to increase risk for dental caries and other adverse dental outcomes) it was difficult to disentangle the effects of the water crisis from the effects of the intervention. Therefore, the impact analysis excludes treatment and matched comparison providers with a business practice location in Flint and the broader Genesee County; however, sensitivity tests (not shown) indicate that results were not sensitive to including or excluding Flint and Genesee County providers.

To attribute a beneficiary to a treatment or a matched comparison provider at baseline or follow-up, respectively, the beneficiary must have received the plurality of his or her primary care from the provider during the provider's baseline or follow-up periods, respectively, according to Medicaid claims data. Beneficiaries included in the impact evaluation could be attributed to a different provider at baseline versus follow-up; however, they must have been attributed to a treatment or comparison provider at both baseline and at follow-up for inclusion in the analysis. Beneficiaries included in the comparison group could have no visits with any treatment provider during the follow-up period to ensure that the training and TA component of the intervention did not affect them. Beneficiaries attributed to a PCP at baseline or follow-up were assigned to the enrollment date of the PCP to whom they were attributed.

For the analytic file used for the regression models, the analysis constructed variables on beneficiary-level characteristics (demographic characteristics and Chronic Illness and Disability Payment System score) as well as quarterly use (outcome) measures using Medicaid claims, encounter, and eligibility data. The outcomes for this evaluation included one core measure (number of outpatient emergency department [ED] visits) and five awardee-specific measures (number of fluoride varnish applications, number of oral health evaluations conducted, number of preventive dental visits, number of restorative procedures, number of dental sealant applications, and ambulatory-sensitive outpatient ED visits for dental caries). The analysis constructed all outcomes from Medicaid claims, encounter, and eligibility data. Specifications used for the outcome measures are available upon request.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows. the provider-level variables used for matching. The table displays the weighted means of baseline characteristics for the 896 treatment primary care providers (PCPs) and the 2,743 comparison PCPs matched for the impact analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable, which was calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group of PCPs). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables include provider's gender; credentials (medical doctor, doctor of osteopathic medicine, nurse practitioner, physician assistant, or registered nurse); specialty (pediatrics, family practice, internal medicine, student, or other primary care specialty); sole proprietorship; county or region of the provider's business practice location; urbanicity of the provider's business practice location; county-level median household income of the provider's business practice location; volume of Medicaid- and Children's Health Insurance Program (CHIP)-enrolled children seen in the baseline year; and service use of Medicaid- and CHIP-enrolled children seen in baseline year. The variables are measured over various specified intervals within the 12 months before the provider's enrollment in the intervention. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

The table also shows the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence whether the difference in the means is statistically significant. The equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. The results are used to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

Table B.1. Baseline characteristics of treatment and matched comparison PCPs for Altarum

Characteristic	Treatment mean	Matched comparison mean	Adjusted difference	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
PCP demographics, %							
Female	59	51	7.5	13	0.15	< 0.01	< 0.01
PCP credentials, specialty, and sole proprietorship, %							
PCP credentials: Medical doctor or doctor of osteopathy	91	91	0.15	< +/-1	0.01	0.89	< 0.01
Specialty: Pediatrics	75	76	-1.2	-1.6	-0.03	0.48	< 0.01
Specialty: Family medicine	11	13	-2.2	-21	-0.07	0.07	< 0.01
Specialty: Internal medicine	11	9.4	2.0	18	0.06	0.09	< 0.01
Specialty: Student	1.2	0.56	0.61	52	0.06	0.10	< 0.01
Specialty: Other specialty	2.4	1.6	0.80	33	0.05	0.16	< 0.01
Sole proprietor	16	19	-2.9	-18	-0.08	0.05	< 0.01
PCP business location, county or region, %							
Kent County	4.2	4.5	-0.38	-9.2	-0.02	0.63	< 0.01
Oakland County	12	11	0.26	2.2	0.01	0.83	< 0.01
Wayne County	22	26	-4.1	-19	-0.10	0.01	< 0.01
Southeast Michigan region	10	12	-1.4	-14	-0.05	0.25	< 0.01
Lansing region	8.2	3.4	4.8	58	0.18	< 0.01	0.02
Macomb County	5.9	5.0	0.89	15	0.04	0.32	< 0.01
Upper Peninsula region	0.43	0.42	0.00	1.1	0.00	0.99	< 0.01
East Central region	12	12	0.02	< +/-1	0.00	0.99	< 0.01
North Lower Peninsula region	7.6	5.8	1.8	23	0.07	0.07	< 0.01
Southwest Michigan region	2.9	3.6	-0.72	-25	-0.04	0.26	< 0.01
Genesee County	8.0	7.7	0.22	2.8	0.01	0.83	< 0.01
West Michigan region	2.7	3.2	-0.50	-18	-0.03	0.45	< 0.01
Kalamazoo County	4.6	5.3	-0.78	-17	-0.04	0.34	< 0.01
PCP business practice location, urbanicity, %							
Urban area	81	84	-3.3	-4.1	-0.08	0.03	< 0.01
Suburban area	17	13	4.1	24	0.11	< 0.01	< 0.01
Rural area	2.4	3.1	-0.77	-32	-0.05	0.22	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean	Matched comparison mean	Adjusted difference	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
PCP business practice location, area-level factors							
Median household income by county	51,392	50,233	1,159	2.3	0.08	0.04	< 0.01
PCP-level volume of Medicaid- and CHIP-enrolled children seen in baseline year							
1 to 4 children, %	0.02	0.03	-0.01	-80	-0.01	0.80	< 0.01
5 to 19 children, %	0.25	0.37	-0.12	-50	-0.02	0.53	< 0.01
20 to 99 children, %	4.4	5.2	-0.79	-18	-0.04	0.34	< 0.01
100 to 499 children, %	30	31	-1.2	-3.8	-0.03	0.51	< 0.01
500 to 999 children, %	35	34	1.5	4.3	0.03	0.38	< 0.01
1,000 or more children, %	30	29	0.54	1.8	0.01	0.76	< 0.01
Number of children	1,097	903	194	18	0.14	< 0.01	< 0.01
PCP-level service use of Medicaid- and CHIP-enrolled children seen in baseline year (%)							
One or more inpatient claims	3.6	3.9	-0.36	-10.0	-0.02	< 0.01	< 0.01
One or more inpatient claims for dental caries	0.01	0.02	0.00	-24	0.00	0.10	< 0.01
One or more ED or observation claims	43	46	-3.1	-7.2	-0.06	< 0.01	< 0.01
One or more ED or observation claims for dental caries	0.23	0.25	-0.02	-8.8	0.00	0.07	< 0.01
One or more restorative procedures	17	19	-1.7	-9.8	-0.04	< 0.01	< 0.01
One or more preventive or diagnostic dental claims	48	52	-3.5	-7.2	-0.07	< 0.01	< 0.01
Patients of children that were fee-for-service	2.9	2.4	0.43	15	0.03	< 0.01	< 0.01
Propensity score	0.01	0.01	0.00	41	0.26	< 0.01	0.57
Number of PCPs	896	2,743					

Source: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

Note: Standardized difference is calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison provider is matched to a treatment provider. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid. Exact matching variables include provider credentials (the provider being an MD or DO vs. any other credentials), provider taxonomy category, county or region of provider's business location.

CHIP = Children's Health Insurance Program; DO = doctor of osteopathic medicine; ED = emergency department; FFS = fee for service; MD = medical doctor; PCP = primary care provider.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Tables C.1 through C.3 display the results from the impact analysis. Table C.1 shows the impact estimates for the primary study population using a beneficiary-level fixed effects model, measured separately over intervention Years 1 and 2. To include beneficiaries, this model required. attributing them to a primary care provider (PCP) at both baseline and follow-up. Table C.2 shows similar results for the subgroup of 45,898 treatment beneficiaries (48 percent) who enrolled within the first 12 months of the program start date versus the 49,046 treatment beneficiaries (52 percent) who enrolled after the first 12 months of the launch date. Table C.3 shows the impact estimates for the complete study population using a provider-level fixed effects model that did not require attributing beneficiaries to a PCP at both baseline and follow-up and therefore includes a more complete subset of participants. The models were estimated for number of services used (per 1,000 beneficiaries). The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. Impact estimates that differ statistically from zero at the .10, .05, and .01 levels, using a two-tailed test, are indicated with one, two, or three asterisks, respectively.

Table C.1. Estimated impact of the MCPP on select Medicaid use measures during a 24-month follow-up period: beneficiary-level fixed effects model

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED or observation visits, per 1,000 beneficiaries[b]					
Baseline year	375	410			
Year 1	313	345	3.5 (3.8)	< 1%	0.36
Year 2	263	287	11** (4.4)	3.7%	0.01
Cumulative	292	321	5.3 (3.6)	1.6%	0.14
Preventive dental visits, per 1,000 beneficiaries					
Baseline year	432	429			
Year 1	464	450	10.0*** (2.8)	2.2%	< 0.01
Year 2	463	443	17*** (3.5)	3.6%	< 0.01
Cumulative	465	448	13*** (2.6)	2.8%	< 0.01
Restorative procedures, per 1,000 beneficiaries					
Baseline year	146	149			
Year 1	158	162	-2.3 (2.8)	-1.5%	0.42
Year 2	159	163	-2.0 (3.2)	-1.2%	0.54
Cumulative	158	162	-2.0 (2.5)	-1.3%	0.43
Dental sealants, per 1,000 beneficiaries					
Baseline year	46	47			
Year 1	38	37	1.9 (1.2)	5.1%	0.12
Year 2	27	28	-0.22 (1.3)	<1%	0.86
Cumulative	34	34	0.89 (1.1)	2.6%	0.41
Oral health evaluations, per 1,000 beneficiaries					
Baseline year	184	182			
Year 1	204	198	3.6*** (1.2)	1.9%	< 0.01
Year 2	210	200	7.0*** (1.5)	3.5%	< 0.01

Table C.1 (*continued*)

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Cumulative	207	199	5.1*** (1.1)	2.6%	< 0.01
Fluoride varnish, per 1,000 beneficiaries					
Baseline year	185	184			
Year 1	232	193	37*** (1.3)	19%	< 0.01
Year 2	207	189	17*** (1.6)	8.4%	< 0.01
Cumulative	223	192	29*** (1.2)	15%	< 0.01
Ambulatory-sensitive ED or observation visits for dental caries, per 1,000 beneficiaries					
Baseline year	1.1	1.2			
Year 1	0.88	1.0	-0.05 (0.19)	-5.1%	0.79
Year 2	0.92	0.82	0.19 (0.23)	23%	0.40
Cumulative	0.87	0.95	0.02 (0.18)	1.7%	0.93
Sample sizes					
Number of beneficiaries, all ages					
Baseline year	82,464	137,176			
Year 1	87,628	116,071			
Year 2	82,004	99,456			
Cumulative	94,944	124,696			
Number of beneficiaries, older than 1 year					
Baseline year	73,805	122,998			
Year 1	78,944	105,126			
Year 2	74,733	91,159			
Cumulative	86,023	113,457			
Number of beneficiaries, older than 6 years					
Baseline year	38,393	68,425			
Year 1	39,254	59,827			
Year 2	37,399	52,151			
Cumulative	43,304	65,004			

Sources: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

Note: Impact estimates for number of visits are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] According to the awardee's theory of action, the intervention was not expected to affect the overall ED visit rate. However, according to its theory of action, the intervention was expected to affect the subset of ambulatory-sensitive ED visits for dental caries.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; MCPP = Michigan Caries Prevention Program; SE = standard error.

Table C.2. Estimated impact of the MCPP on select Medicaid use measures during a 24-month follow-up period, for beneficiaries enrolled within 12 months of program start date: beneficiary-level fixed effects model

	Beneficiaries enrolled within 12 months of program start date					Beneficiaries enrolled after 12 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED or observation visits, per 1,000 beneficiaries[b]										
Baseline year	381	445				367	363			
Year 1	323	373	14** (5.8)	4.3%	0.01	302	304	-6.8 (5.1)	-1.9%	0.18
Year 2	274	313	26*** (5.8)	9.6%	< 0.01	246	256	-14** (6.8)	-4.3%	0.04
Cumulative	297	343	19*** (5.1)	6.4%	< 0.01	284	290	-10** (4.9)	-2.8%	0.04
Preventive dental visits, per 1,000 beneficiaries										
Baseline year	427	422				439	438			
Year 1	476	452	20*** (4.1)	4.2%	< 0.01	453	452	-0.20 (3.9)	< 1%	0.96
Year 2	470	445	20*** (4.4)	4.4%	< 0.01	451	439	12** (6.0)	2.7%	0.04
Cumulative	474	449	20*** (3.6)	4.3%	< 0.01	453	449	3.3 (3.7)	< 1%	0.37
Restorative procedures, per 1,000 beneficiaries										
Baseline year	144	141				149	159			
Year 1	159	159	-3.5 (4.1)	-2.2%	0.40	156	168	-1.5 (3.7)	-1.1%	0.68
Year 2	160	161	-4.1 (4.1)	-2.5%	0.32	155	165	1.2 (5.1)	< 1%	0.81
Cumulative	160	160	-3.6 (3.5)	-2.3%	0.31	156	167	-0.49 (3.5)	< 1%	0.89
Dental sealants, per 1,000 beneficiaries										
Baseline year	45	45				47	49			
Year 1	40	38	1.4 (1.8)	3.7%	0.42	37	37	2.0 (1.6)	5.7%	0.21
Year 2	27	27	-0.20 (1.7)	< 1%	0.90	29	31	-0.48 (2.1)	-1.5%	0.82
Cumulative	33	33	0.47 (1.5)	1.4%	0.76	35	35	1.3 (1.5)	3.8%	0.39
Oral health evaluations, per 1,000 beneficiaries										
Baseline year	186	179				182	184			
Year 1	210	196	6.9*** (1.8)	3.4%	< 0.01	198	200	0.49 (1.7)	< 1%	0.77
Year 2	214	202	5.5*** (1.9)	2.6%	< 0.01	205	196	11*** (2.6)	6.0%	< 0.01
Cumulative	212	199	6.1*** (1.6)	3.0%	< 0.01	201	199	3.7** (1.6)	2.0%	0.02

Table C.2 (*continued*)

	Beneficiaries enrolled within 12 months of program start date					Beneficiaries enrolled after 12 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Fluoride varnish, per 1,000 beneficiaries										
Baseline year	183	182				188	186			
Year 1	239	196	42*** (1.9)	21%	< 0.01	226	191	32*** (1.8)	17%	< 0.01
Year 2	210	192	17*** (2.0)	8.4%	< 0.01	204	184	18*** (2.6)	9.4%	< 0.01
Cumulative	225	194	29*** (1.7)	15%	< 0.01	221	190	29*** (1.7)	15%	< 0.01
Ambulatory-sensitive ED or observation visits for dental caries, per 1,000 beneficiaries										
Baseline year	1.2	1.3				0.97	1.1			
Year 1	1.0	1.0	0.03 (0.30)	3.3%	0.91	0.72	0.99	-0.12 (0.25)	-12%	0.64
Year 2	0.80	0.88	-0.03 (0.29)	-3.8%	0.91	1.3	0.73	0.70* (0.39)	96%	0.07
Cumulative	0.91	0.97	-0.02 (0.26)	-1.9%	0.94	0.83	0.92	0.06 (0.24)	6.5%	0.79
Sample sizes										
Number of beneficiaries, all ages										
Baseline year	45,195	71,135				37,269	66,041			
Year 1	40,926	45,934				46,702	70,137			
Year 2	42,767	46,647				39,237	52,809			
Cumulative	45,898	50,807				49,046	73,889			
Number of beneficiaries, older than 1 year										
Baseline year	40,871	64,286				32,934	58,712			
Year 1	36,938	41,496				42,006	63,630			
Year 2	39,120	42,630				35,613	48,529			
Cumulative	41,773	46,186				44,250	67,271			
Number of beneficiaries, older than 6 years										
Baseline year	20,849	34,758				17,544	33,667			
Year 1	18,217	23,352				21,037	36,475			
Year 2	19,594	24,078				17,805	28,073			
Cumulative	21,028	26,260				22,276	38,744			

Table C.2 (*continued*)

Sources: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

Note: Impact estimates for number of visits are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] According to the awardee's theory of action, the intervention was not expected to affect the overall ED visit rate. However, according to its theory of action, the intervention was expected to affect the subset of ambulatory-sensitive ED visits for dental caries.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; MCPP = Michigan Caries Prevention Program; SE = standard error.

Table C.3. Estimated impact of the MCPP on select Medicaid use measures during a 24-month follow-up period: provider-level fixed effects model

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED or observation visits, per 1,000 beneficiaries[b]					
Baseline year	351	397			
Year 1	339	359	26*** (6.0)	8.0%	< 0.01
Year 2	281	279	48*** (8.9)	20%	< 0.01
Cumulative	315	327	34*** (6.4)	12%	< 0.01
Preventive dental visits, per 1,000 beneficiaries					
Baseline year	414	405			
Year 1	458	442	6.7 (4.2)	1.5%	0.11
Year 2	460	438	13** (5.9)	2.8%	0.03
Cumulative	461	442	9.7** (4.5)	2.2%	0.03
Restorative procedures, per 1,000 beneficiaries					
Baseline year	144	145			
Year 1	160	162	-1.2 (2.6)	< 1%	0.65
Year 2	162	164	-1.8 (3.2)	-1.2%	0.56
Cumulative	161	164	-1.7 (2.5)	-1.1%	0.49
Dental sealants, per 1,000 beneficiaries					
Baseline year	43	42			
Year 1	36	35	-0.28 (1.2)	< 1%	0.82
Year 2	26	26	-0.21 (1.4)	< 1%	0.88
Cumulative	32	32	-0.35 (1.1)	< 1%	0.76
Oral health evaluations, per 1,000 beneficiaries					
Baseline year	179	175			
Year 1	203	195	4.2** (1.8)	2.2%	0.02
Year 2	208	198	6.3** (2.8)	3.2%	0.02
Cumulative	206	197	5.3** (2.1)	2.7%	0.01
Fluoride varnish, per 1,000 beneficiaries					
Baseline year	176	171			
Year 1	224	186	33**** (3.3)	18%	< 0.01
Year 2	205	185	15**** (3.0)	8.0%	< 0.01
Cumulative	218	187	26**** (2.7)	14%	< 0.01
Ambulatory-sensitive ED or observation visits for dental caries, per 1,000 beneficiaries					
Baseline year	1.1	1.3			
Year 1	1.1	1.3	-0.09 (0.18)	-7.5%	0.64
Year 2	1.1	0.93	0.30 (0.20)	41%	0.12
Cumulative	1.1	1.2	0.07 (0.17)	6.8%	0.68

Table C.3 (*continued*)

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes					
Number of beneficiaries, all ages					
Baseline year	126,191	242,770			
Year 1	120,993	196,181			
Year 2	119,403	178,813			
Cumulative	137,137	221,070			
Number of beneficiaries, older than 1 year					
Baseline year	115,072	222,240			
Year 1	109,608	179,379			
Year 2	109,439	165,199			
Cumulative	124,973	203,031			
Number of beneficiaries, older than 6 years					
Baseline year	65,112	135,802			
Year 1	58,072	108,298			
Year 2	58,199	99,912			
Cumulative	66,998	123,578			

Sources: Mathematica's analysis of Medicaid beneficiaries attributed to providers enrolling in the intervention through May 31, 2017, and Medicaid claims and eligibility data obtained from the Michigan Department of Health and Human Services from May 1, 2014, through December 31, 2017.

Note: Impact estimates for number of visits are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] According to the awardee's theory of action, the intervention was not expected to affect the overall ED visit rate. However, according to its theory of action, the intervention was expected to affect the subset of ambulatory-sensitive ED visits for dental caries.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; MCPP = Michigan Caries Prevention Program; NPI = National Provider Identifier; SE = standard error.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: American College of Cardiology Foundation

September 2020

Anna Hill, Dana Jean-Baptiste, Shannon Heitkamp, and Randall Brown

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION

The American College of Cardiology Foundation (ACCF) used funding from Round 2 of the Health Care Innovation Awards (HCIA R2) to create the SMARTCare program. The goal of the program was to improve the appropriateness of care and quality of life for stable ischemic heart disease (SIHD) patients and to increase their adherence to coronary artery disease (CAD) treatment guidelines. The key innovations within SMARTCare's design were to combine four related decision-support tools that used evidence-based medicine, plus a bundled approach for paying for these services. The program sought to engage all patients with SIHD treated by participating physicians. The program launched in November 2014 and the intervention period covered under HCIA R2 ended in February 2018, after a six-month no-cost extension. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program included four decision-support tools that used evidence-based medicine to improve care for Medicare beneficiaries with SIHD.
- Because the program selected and enrolled patients based on clinical judgment and criteria using measures not available in claims, it was not possible to identify a credible comparison group. As a result, it was not possible to estimate the impact of the SMARTCare intervention on service use and costs.
- This report presents the demographic and health characteristics at enrollment for the 2,455 Medicare FFS beneficiaries who participated in SMARTCare and met the claims-based criteria for inclusion in the analysis.

The program relied on health information technology (health IT) to (1) provide clinical decision support for managing SIHD to cardiologists and other clinical specialists at the point of care, (2) support patient–clinician shared decision making, and (3) enable the use of clinical registries to track and improve care. The awardee expected that the tools would guide clinicians' decisions—from ordering tests to performing procedures—and, in turn, reduce inappropriate use of cardiac screening tests and procedures and reduce rates of SIHD-related complications. The health IT tools also provided customized, patient-specific estimates of the risks and benefits of specific procedures, as well as educational materials to support shared decision making.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	ACCF focused on changing clinicians' behavior by providing (1) decision-support tools at the point of care to assess treatment options for SIHD, (2) patient education materials on specific treatment options, and (3) individually tailored risk and benefit information to support shared decision making.
Major innovation	ACCF's major innovations were the use of health IT to bundle evidence-based, decision-support tools, as well as the awardee's bundled approach to paying for these services.

Table 1 (*continued*)

Program characteristics	Description
Program components	• Two EMR-based tools to aid in clinical decision support • Two EMR-based tools to aid in shared patient decision making • Tools to aid in enrolling patients and tracking outcomes
Target population	The target population was patients with SIHD.
Total enrollment	The awardee enrolled 29,053 patients in the program (more than 100 percent of the original enrollment goal); less than 10 percent of participants were enrolled in FFS Medicare.
Level of engagement	The awardee did not track patient-level engagement. Engagement at the provider level varied, and a few participating providers drove program enrollment.
Theory of change or theory of action	Improving risk communication and shared decision making between participants and cardiac clinicians would lead to optimizing medication and adjusting lifestyle factors for the greatest potential impact on a participant's risk factors.
Award amount	$15,830,092
Effective launch date	The SMARTCare program began operating in November 2014.
Program settings	Provider-based settings, including primary and specialty care clinics, hospitals, and academic medical centers
Market area	Rural, urban, and suburban (Florida and Wisconsin)
Target outcomes	The program sought to improve SIHD patients' care and health outcomes, and did not expect to have large effects during the demonstration period on expenditures or use of acute care. Key outcomes for the program were: • Decrease in the risk-adjusted bleeding complication rate for elective PCIs • Improvement in either the Seattle Angina Questionnaire score (patients with chest pain) or the Heart Quality of Life score (patients without chest pain) • Increase in adherence to CAD treatment guidelines, including – Angiotensin converting enzyme inhibitor or angiotensin receptor blockers therapy prescribed for participants with diabetes or LVSD – Oral antiplatelet therapy prescribed for participants with CAD – Aspirin or other antithrombotic prescribed for participants with acute myocardial infarction, coronary artery bypass graft, PCI, or LVSD – Lipid control prescribed for participants with CAD – Beta-blocker therapy prescribed for participants with CAD – Tobacco use assessment and tobacco cessation counseling administered to participants
Payment model	Value-based bundled payments
Sustainability plans	ACCF reported plans to sustain the program by simplifying it, training implementing sites to operate the program independently, and developing the payment model. Aspects of the program that the awardee will sustain include ePRISM, clinical decision support, and the program's general effect on the site's strategic planning. The awardee also reported creating program resources to help sites sustain aspects of the program they find valuable, such as a video explaining the program and a "SMARTCare 101" fact sheet.

ACCF = American College of Cardiology Foundation; CAD = coronary artery disease; CMS = Centers for Medicare & Medicaid Services; E&M = evaluation and management; EMR = electronic medical record; FFS = fee-for-service; HCC = Hierarchical Condition Category; IT = information technology; LVSD = left ventricular systolic dysfunction; PCI = percutaneous coronary intervention; SIHD = stable ischemic heart disease.

Because the program selected and enrolled patients based on clinical judgment and criteria using measures not available in medical claims, it was not possible to construct a comparison group of patients matched to those enrolled in the intervention. An alternative approach—obtaining unbiased estimates by comparing all those meeting eligibility criteria assessable in claims to a similar comparison group—also was not feasible. Only 3 percent of beneficiaries who were seen by participating providers and met the eligibility criteria assessable with Medicare claims actually participated in the program. Thus, it would have been highly unlikely that statistically significant impact estimates would be observed in the sample of eligibles even if the true impacts of the program on the subset who actually participated were very large. As a result, it was not possible to obtain reliable estimates of the SMARTCare intervention's impact on outcomes. Instead, the descriptive analysis presented in this report was limited to presenting the baseline characteristics of 2,455 Medicare FFS beneficiaries who participated in SMARTCare and met the claims-based criteria for including them in the analysis. Table 2 summarizes the key features of the descriptive analysis.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	A rigorous impact evaluation of this program was not possible, primarily because enrollment decisions were based on clinician judgement, and could not be replicated in Medicare claims data. An analysis using all eligible Medicare beneficiaries as the treatment group would have been unbiased but was not feasible due to the low participation rate among beneficiaries meeting the eligibility criteria assessable with claims.
Intervention group for descriptive analysis	The descriptive analysis included 2,455 Medicare FFS beneficiaries, representing 44 percent of all Medicare participants listed in the awardee's finder file (and 8 percent of all participants). The study excluded 2,502 Medicare beneficiaries who were enrolled in Medicare Advanaged, 523 who were not enrolled in both Medicare Parts A and B, 90 for whom Medicare was not their primary payer, and 50 with fewer than 90 days of claims history before enrollment. Appendix A, Table A.1 describes the identification of the analytic sample.
Limitations	Due to the problems noted above, the findings presented in this report cannot be used to make inferences about the impact of SmartCare on Medicare service use and spending or other program outcomes.

ECC = early childhood caries.

PROGRAM DESIGN AND ADAPTATION

The SMARTCare program relied on a suite of health IT tools aimed at improving care for patients with SIHD.[1] The health IT tools (1) provided clinical decision support, (2) aided in shared physician-patient decision making, and (3) assisted in patient enrollment and tracking outcomes.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Health IT tools to provided clinical decision support

The SMARTCare program included two health IT tools intended to provide clinical decision support. FOCUS is a computerized decision support tool that incorporates participant-specific information to determine whether ordered imaging meets appropriate use criteria and, if so, which test is most appropriate for a specific participant. IndiGO calculates and displays the personalized risk of an adverse event and suggests and prioritizes approaches with the greatest potential to reduce that risk.

Health IT tools to aided in shared patient decision making

The program also included two tools to aid in shared decision making. ePRISM is a shared decision-making tool that produces a customized, patient-specific consent form. The consent form provides participant education along with individualized estimates of benefits and risks of complications tailored to each participant before invasive cardiac catheterization or PCI. eLumen uses patient information from ePRISM to recommend actions that physicians can take to reduce complications during or following the procedure. Patient education materials developed by Health Dialog inform and prepare patients for more effective conversations with their physicians about their conditions and treatment options.

Health IT tools to assisted in patient enrollment and tracking outcomes

The SMARTCare intervention also included a fifth EMR-based tool, Tonic. Tonic is a web-based application that runs on an iPad and a participant can use it to consent to the collection of his or her health data and to provide participant-reported outcomes before and after treatment. In addition to EMR-based tools, practices used two registries to establish a national benchmark against which outcomes for individual providers, health systems, or entire states can be compared.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

ACCF was partly successful in implementing its program. The awardee enrolled 29,053 participants, more than 100 percent of its original enrollment target. However, a relatively few clinicians were responsible for the awardee's enrollment success. There was a general lack of clinician buy-in across all sites. The widespread lack of clinician buy-in stemmed primarily from two factors: (1) difficulty using the tools because of interoperability issues within sites' electronic medical record (EMR) systems, and (2) skepticism among clinicians that the tools would improve patients' care. In addition, several operational issues prevented the awardee from implementing the program with full success. For example, adverse payment incentives for clinicians operating in a fee-for-service (FFS) environment worked against the program's goal of reducing inappropriate testing, which led to reduced revenue from patients with FFS coverage. In addition, interoperability issues apart from the EMR compatibility issue mentioned earlier delayed the delivery of services and led to variation in how sites used the tools. Some sites used

the tools through a web-based system. Others created manual, paper-based workarounds instead of using the FOCUS tool at the point of care.

Although the awardee was unable to deliver services as originally designed, participating clinicians reported that they felt the program had a positive effect on care delivery. They also offered positive feedback about some of the SMARTCare tools. However, participating clinicians questioned the feasibility of effectively working with multiple tools at the same time. One interview respondent noted that the tools were more efficient than previous procedures to address the appropriateness of test ordering, but said that the tools would be more successful if they had been integrated into one individual tool instead of separate and disparate tools. The FOCUS tool used to assess patients' risk appeared to be the most useful tool, with 70 percent of clinician survey respondents reporting that the program improved their ability to assess whether diagnostic tests for a patient met the criteria for appropriate use.

Implications of program implementation for achieving program goals

- The awardee met its enrollment target, but a small group of providers accounted for the majority of enrollees.
- Only one site implemented the entire suite of SMARTCare tools; all other sites were unable to use all of the tools. Providers also reported that working with multiple tools for each patient was cumbersome.
- Lack of widespread buy-in among participating clinians, due in part to incentives under Medicare FFS to order tests, might have limited the effectiveness of the intervention.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Study sample

The descriptive analysis included 2,455 Medicare FFS beneficiaries, representing 44 percent of the 5,620 Medicare participants listed in the awardee's finder file (and 8 percent of the total 29,053 enrollees). (Most participants in SmartCare were commercially insured; usable data for the Medicaid participants could not be obtained for this study.) Of the 5,620 Medicare participants, the study excluded 2,502 who were enrolled in Medicare Advantage, 523 who were not enrolled in both Medicare Parts A and B, 90 for whom Medicare was not their primary payer, and 50 with fewer than 90 days of claims history before enrollment. Appendix A, Table A.1 describes the identification of the sample used for the descriptive analysis.

Characteristics of Medicare FFS participants

As expected given the focus of the intervention, Medicare FFS participants appear to be somewhat sicker at enrollment than Medicare FFS beneficiaries nationally. The average hierarchical condition category risk score among Medicare FFS participants was 1.4, indicating

that their predicted total Medicare expenditures were 40 percent higher than the average for Medicare FFS beneficiaries nationally. However, slightly fewer (11 percent) than the national average of 18 percent were dually eligible for Medicaid. Also consistent with the program's target population, nearly one-half (49 percent) of the Medicare FFS participants had a history of chest pain and nearly one-quarter (22 percent) had congestive heart failure during the year before enrollment in SmartCare. Other common chronic conditions among participants were diabetes with and without complications (37 percent) and vascular disease (24 percent).

SMARTCare's Medicare FFS participants were also more likely to use emergency and acute inpatient services during the year before enrollment than Medicare FFS beneficiaries nationally in 2017. During the baseline year, Medicare participants had on average 315 hospitalizations and 613 ED visits per 1,000 beneficiaries, exceeding the national averages by 30 and 55 percent, respectively. Average total Medicare spending among participants was $973 per beneficiary per month during the baseline year, about 16 percent higher than the national Medicare FFS average of $840. The SMARTCare program was designed to improve care for cardiac patients with SIHD, therefore slightly higher than average rates of ED and inpatient service use and spending would be expected.

The intervention primarily aimed to improve the appropriate use of cardiac imaging procedures as well as PCI, which might lead to reductions in the use of these procedures. Unsurprisngly, the vast majority of the Medicare FFS SMARTCare participants had a cardiac testing procedure in the 12 months before enrollment in the intervention: 95 percent had a cardiology diagnostic lab test, 71 percent had an electrocardiogram, and 53 percent has a diagnostic imaging procedure. Interventions like PCI and stent placement were rare in the year before enrollment, indicating that most of the Medicare FFS patients enrolled in the program had stable heart disease.

Table 3. Baseline characteristics of Medicare FFS participants

Characteristic	Participants (N = 2, 554)
Demographics	
Age at enrollment, years	72
Age group, %	
Younger than 65	10
65 to 74	53
75 to 84	30
85 and older	7
Female, %	46
White, %	86
Black, %	8
Other, %	6
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	80
Disability insurance benefits	18
End stage renal disease	2

Table 3 (*continued*)

Characteristic	Participants (N = 2, 554)
Medicare-Medicaid dual eligibility status, %	
Dually eligible for Medicare and Medicaid	11
HCC score[a]	
Mean	1.4
25th percentile	0.7
Median	1.1
75th percentile	1.8
Health status and diagnosis, %	
Chest pain	49
COPD	16
Congestive heart failure	22
Diabetes without complications	14
Diabetes with complications	23
Major depressive, bipolar, and paranoid disorders	7
Morbid obesity	7
Vascular disease	24
Service use and expenditures during the year before enrollment	
Total Medicare expenditures ($ PBPM)	973
Hospital admissions, per 1,000 beneficiaries	315
ED outpatient visits, per 1,000 beneficiaries	613
Any hospitalization, %	21
Any ED visit, %	34
Diagnostic imaging and cardiac intervention procedures during the year before enrollment, %	
Any electrocardiogram	71
Any diagnostic imaging	53
Any cardiology diagnostic lab test	95
Any stent placement	1.0
Any PCI	0

Source: Mathematica's analysis of Medicare claims and enrollment data from March 2014 through August 2017.

Note: The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

COPD = chronic obstructive pulmonary disease; ED = emergency department; FFS = fee for service; HCC = hierarchical condition category; PBPM = per beneficiary per month; PCI = peripheral component interconnect.

Challenges of measuring program impacts

It was not possible to obtain credible estimates of program impacts. The criteria that clinicians used to identify and enroll participants into the program were not observable for potential comparison cases and were likely to affect patients' outcomes. Thus, it was not possible to select an equivalent comparison group to obtain unbiased impact estimates. Estimating impacts over all eligibles to obtain unbiased estimates was not a viable alternative, because only 3 percent of

Medicare FFS beneficiaries who met the eligibility criteria that could be assessed with Medicare claims participated in the program. Thus, it would have been highly unlikely for the evaluation to find statistically significant estimates of program effects if the study estimated impacts over all eligible beneficiaries, even if the true program impacts had been quite large.

CONCLUSION

ACCF partly succeeded in implementing the SMARTCare program to improve the appropriateness of care and quality of life for patients with SIHD and to increase adherence to CAD treatment guidelines. Enrollment was successful, with 29,053 participants—108 percent of its original enrollment target—by the end of the initial cooperative agreement. However, most enrollees were patients of a relatively small number of clinicians, due to lack of widespread clinician buy-in. Many operational issues also prevented the program from being fully successful, including EMR interoperability problems with the SMARTCare tools and adverse payment incentives for clinicians operating in a FFS environment—a reduction in imaging and PCI procedures could substantially reduce reimbursement levels for clinicians and facilities.

PROGRAM SUSTAINABILITY

By the end of the ACCF award in July 2018, the SMARTCare program ceased operating as originally implemented due to lack of funding, implementation challenges, and inconsistent provider participation. None of the participating sites continued using the main suite of SMARTCare tools, although some sites retained minor program components.

ACCF proposed a bundled payment to support diagnosing and treating patients with SIHD. The awardee faced three main challenges to advancing its proposed model. First, only one of the nine participating sites fully implemented the program. This was largely due to EMR interoperability problems with the SMARTCare tools and lack of provider buy-in, both of which are threats to program sustainability moving forward. Second, the awardee had difficulty accessing the claims data needed to determine the underlying costs of the program and therefore could not arrive at an appropriate payment amount. Third, the awardee and its payer partner were

ACCF's proposed payment model

ACCF proposed a bundled payment to support diagnosing and treating patients with SIHD. The payment would cover (1) all E&M services by cardiologists for one month following an initial patient visit to a physician for new or significantly changed angina symptoms; and (2) stress tests, angiograms, and angioplasties during the six months before treatment began for current symptoms.

To trigger the payment, a provider would use the program's electronic decision support tool to identify eligible patients and then submit an E&M code to the payer. The payer would classify the patient into one of three levels based on appropriateness of testing, test results, and quality of outcomes. The payer would reimburse the provider based on the difference between actual spending for the patients in a group and the projected budget for the group.

unsuccessful in setting the trigger for the bundled payment. As an alternative, the awardee and payer considered providing hospitals and health systems a care coordination fee in exchange for their data on outcomes and complications. However, negotiations to set the fee did not occur and the awardee did not expect fees to sustain the program.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Among SmartCare's 29,053 enrollees, only 5,620 were Medicare beneficiaries. About half of these individuals were in Medicare Advantage programs, for which claims data were not available. The descriptive analysis in this report relied on the 2,455 participants enrolled in Medicare fee-for-service and who met the other standard claims-basead study inclusion criteria. Table A.1 shows the identification of the analytic sample.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total Medicare beneficiaries as of September 2017		**5,620**
Missing enrollment date	0	5,620
Did not meet study's standard claims-based inclusion criteria		
Not enrolled in both Part A and B	523	5,097
Enrolled in Medicare Advantage	2,502	2,595
Medicare not primary payer	90	2,505
Fewer than 90 days of claims history before enrollment	50	2,455
Final Medicare FFS beneficiaries in descriptive analysis		**2,455**

Source: Mathematica's analysis of awardee's finder file and Medicare claims and enrollment data from March 2015 through August 2017, as of September 2017.

FFS = fee-for-service.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Amerigroup

Setptember 2020

Lisa M. Lines, Grace Anglin, Michaella Morzuch, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

AMERIGROUP

Amerigroup, the sole Medicaid managed care provider for Georgia's foster care program, received a Round 2 Health Care Innovations Award (HCIA R2) to implement the Coaching and Comprehensive Health Supports (COACHES) program. Through COACHES, Amerigroup and its partner, Families First, provided intensive health coaching services for youth who lived in participating counties in Georgia and were about to transition out of foster care. Youth in foster care often experience significant mental and emotional stress during childhood and lack a continuous adult support system. As a result, they are often ill equipped for the transition to adulthood. The COACHES program paired participating youth with a health coach employed and trained by Families First to help prepare them for this transition by educating them about the health care and social services systems, helping them build life skills, and supporting them as they advocated for their own needs. Table 1 summarizes the program's key characteristics.

Amerigroup expected that, by working with a coach, youth would experience an increase in health literacy, the use of primary care and preventive services, educational attainment and employment, connections to peer and adult social supports, life skills, and knowledge of legal justice systems. As a result, youth would experience improved service coordination, which the awardee expected to eventually reduce the total cost of care.

Important issues for understanding the evaluation

- The COACHES program provided intensive health coaching services to youth in 34 counties in Georgia who were about to transition out of foster care.
- Amerigroup theorized that working with health coaches would increase health literacy, service coordination, the use of primary care and preventive servives, educational attainment, employment, life skills, and knowledge of the legal and justice systems. Many of these outcomes are not observable in the claims data.
- The awardee expected improved service coordination to eventually lead to an estimated reduction in the cost of care.
- The youth who volunteered to participate self-referred into the program, so they might have differed from youth in the comparison group in ways (such as motivation and ability to manage their own health care) that Medicaid claims data cannot capture.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The COACHES program connected youth who were about to transition out of foster care with an employed, trained health coach who taught them how to access, coordinate, and manage health and social services on their own.
Major innovation	Unlike many other programs for youth in foster care, Families First made participation in COACHES voluntary and participant-driven. For example, participants determined how often they met with their health coaches and the focus of their work in the program.
Program components	• Patient and family engagement • Care management services • Outpatient care coordination
Target population	The program sought to engage youth ages 17 to 20 who had been in foster care for 12 months or longer, had a documented history of behavioral health needs, and resided in one of 34 counties in Georgia in the program's catchment area. Youth who resided with foster families or in group homes had to get permission from their foster parents or group home staff to participate in the program.
Participating providers	Families First, a nonprofit provider of foster care services, delivered the program to youth. Families First provided COACHES services in three office locations (Atlanta, Macon, and Columbus).
Total enrollment	The awardee enrolled 860 patients in the program from September 2014 through August 2017, representing 119 percent of its original enrollment target.
Level of engagement	The median length of time that coaches worked with youth was 3 months, considerably less than the 12 to 18 months the awardee had planned. Engagement was more successful in rural areas.
Theory of change or theory of action	Amerigroup hypothesized that youth who worked closely with a coach would better understand what services they needed and how to access them. Participants would then increase use of primary care, pregnancy prevention services, and educational and employment programs, which in turn would result in better health and social outcomes, as well as lower health and social service costs.
Award amount	$5,833,492
Effective launch date	March 1, 2015
Program settings	Community- and home-based settings
Market area	Urban, rural, and suburban; participating counties in Georgia (Baldwin, Bartow, Bibb, Carroll, Cherokee, Clayton, Cobb, Coweta, Dawson, DeKalb, Douglas, Fayette, Forsyth, Fulton, Gwinnett, Hall, Hancock, Harris, Henry, Houston, Jones, Macon, Newton, Muscogee, Paulding, Peach, Pike, Polk, Randolph, Rockdale, Spalding, Steward, Talbot, and Taylor)
Target outcomes	• Improved health literacy and ability to navigate the health care system • Increased use of primary care and preventive services • Higher educational attainment and increased employment • Improved connections to peer and adult social supports, and life skills (including renting an apartment and household budgeting) • Increased knowledge of legal and juvenile justice systems • Decreased health and social service costs
Payment model	Amerigroup did not develop a payment model because its partner, Families First, secured state funding to sustain the program.

Table 1 (*continued*)

Program characteristics	Description
Sustainability plans	Families First received a contract from the state of Georgia to sustain the program for about one year following the award period. Families First planned to use private donations and scale-back the program after state funding ended.

COACHES = Coaching and Comprehensive Health Supports.

Of the 860 youth enrolled in the program, the treatment group for this evaluation includes 299 youth, excluding those who did not meet the eligibility criteria for the program (age 17 to 20, enrolled in foster care for 12 months or longer, had documented behavioral health needs, and resided in 1 of 34 counties in Georgia in the program's catchment area). The comparison group consists of 570 youth in foster care who met the same criteria, except that they resided in Georgia counties where Amerigroup did not implement the COACHES program. Table 2 summarizes the key features of the evaluation. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries before versus after enrollment relative to the change in outcomes among a matched comparison group.
Intervention group for evaluation	The impact analysis included 299 of the 860 youth who enrolled in the COACHES program, representing 35 percent of total enrollment. The analysis excluded participants if they did not provide HIPAA consent to participate in the research study (298), lacked unique Medicaid identifiers or sufficient Medicaid data (29), or did not meet the program eligibility criteria that could be replicated in the Medicaid claims data (234).
Comparison group	The matched comparison group consisted of 573 children youth in foster care who met the COACHES eligibility criteria, but resided in counties in Georgia where Amerigroup did not implement the COACHES program.
Limitations	The impact evaluation has three main limitations. First, the impact estimates are likely biased by selection because participants self-referred into the program, though the direction and size of the bias is unclear. Second, although the treatment group included youth who resided in the Atlanta metropolitan area (as well as several rural counties), there was not a similarly sized urban area in Georgia from which to draw the comparison group. Third, the final sample sizes were likely too small to detect program effects on most measures unless the true effects were quite large.

COACHES = Coaching and Comprehensive Health Supports; HIPAA = Health Insurance Portability and Accountability Act of 1996.

PROGRAM DESIGN AND ADAPTATION

The program included three main components: (1) patient and family engagement, (2) care management, and (3) outpatient care coordination.[1]

Patient and family engagement

The program was youth directed, which was intended to help engage participants. This meant that the youth determined the frequency with which they met with their coaches and the focus of their work together. Through their meetings, coaches educated youth about the health care and social services systems, helped them build life skills (such as how to rent an apartment and how to construct and live within a household budget), and supported them as they advocated for their own needs. In the second program year, coaches started hosting group education sessions with participants.

Care management

Coaches were not responsible for managing health and social services for participants, but they were responsible for helping participants manage those services. When a youth first enrolled in the program, a coach completed a series of standardized psychosocial and trauma assessments to better understand his or her strengths and needs. The youth and coach then worked together to develop and implement a coaching skills plan that set out the steps the youth could take to meet the goals, including referrals to community-based services and supports (such as the youth's Amerigroup health care coordinator, employment and education services, or weight loss programs). Coaches asked youth to set at least one personal goal in each of the program's five focus areas: (1) education, (2) employment, (3) mental health, (4) physical health, and (5) pregnancy prevention. Amerigroup anticipated the health coaches and youth would work together for 12 to 18 months to implement these plans and achieve their goals.

Outpatient care coordination

Coaches helped youth coordinate the services they received from various medical and behavioral health providers, child service agencies, and community organizations. To do so, coaches aimed to meet regularly with the youth, their service providers, and informal supports (such as religious leaders or family members). The program initially intended to establish its own meetings. However, to avoid duplicating services, staff decided to coordinate with Department of Family and Child Services staff to attend their established family team meetings. To encourage collaboration, program leaders frequently reached out to various service providers to educate them about the program, seek their feedback on implementation, and encourage their involvement in care coordination.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

By the end of the third year, Amerigroup exceeded its enrollment targets, enrolling 860 participants, or about 119 percent of its final three-year projections. However, to reach its enrollment target, Amerigroup had to adjust its initial recruitment strategy. In the first program year, the program started allowing youth to self-refer to the program, in addition to receiving third-party referrals from caseworkers and foster care providers. It also expanded its catchment area and eligibility criteria. Specifically, Amerigroup expanded from 6 to 11 counties in Year 1, from 11 to 18 in Year 2, and from 18 to 34 in Year 3. In addition, Amerigroup expanded from serving only youth who lived in group homes to include youth enrolled in independent living programs early in Year 1, and to include youth who lived with foster families late in Year 1. The program ultimately enrolled about 30 percent of the transition-age youth in foster care in the 34 implementation counties with a documented mental health condition in Medicaid data.

Implications of implementation for detecting program impacts

Program staff thought the COACHES program had a positive impact on participants. However, several factors could have limited the impact of the program on health-related outcomes:

- The awardee planned that participants would enroll for 12 to 18 months, but the median length of time that participants were in the program was 3 months.
- Youth focused most on education- and employment-related goals, not health-related goals.
- A large proportion of participants self-referred into the program, and some coaches thought the program enrolled relatively high-functioning youth as a result. These youth might have had fewer needs and lower spending levels at baseline compared with youth referred by foster care providers, meaning there was less opportunity for improvement as a result of the program.

The awardee successfully hired, trained, and retained the number of coaches it needed to implement the program. Before they started working with the youth, coaches received about 100 hours of in-person training on evidence-based strategies, including techniques for motivational interviewing and tools based on the Transition to Independence Process Model. Staff consistently reported using the evidence-based practices integrated into the COACHES model—most notably motivational interviewing—to encourage participants to manage their health and social services. Moreover, awardee leaders, drawing on data from focus groups they conducted with participants, reported that most coaches followed established implementation protocols, adhering to the model in terms of the frequency and type of interactions between coaches and youth, and maintaining the youth-driven nature of the program.

However, participants spent only a short time in the program. Their median time in the COACHES program was 3 months, considerably less than the 12 to 18 months the awardee had

expected. During their short time in the program, most youth focused on goals related to education or employment. In the final program year, to account for a shorter-than-expected enrollment period, the awardee reduced the number of initial assessments and the frequency of all assessments. The awardee also added the option for youth in foster care who disenrolled from the full COACHES program to continue periodic check-ins with their coaches.

ESTIMATING PROGRAM IMPACTS

Study sample

This impact analysis included 299 (35 percent) of the 860 youth who enrolled in the COACHES program. Participants were excluded who did not provide Health Insurance Portability and Accountability Act of 1996 (HIPAA) consent to participate in the research study, lacked unique Medicaid identifiers or sufficient Medicaid data, or did not meet the program eligibility criteria (age 17 to 20; continuous enrollment in foster care during the year pror to enrollment, documented behavioral health need, and resided in one of 34 counties in Georgia) that could be replicated in the Medicaid claims data. Appendix A, Table A.1 provides additional details on identifying the analysis sample.

A matched comparison group was constructed that met all of the program eligibility criteria except that they resided in Georgia counties where Amerigroup did not implement the COACHES program. To put into practice the requirement that beneficiaries have a documented history of behavioral health needs, the analysis included both treatment and comparison group beneficiaries only if they had mental health or substance use claims in the baseline period. (There were 28 participants excluded for this reason; these beneficiaries might have had behavioral health needs documented in other ways—for example, in their foster care case files—and the analysis excluded them because similar data were not available for the comparison group.)

Characteristics of treatment and comparison group beneficiaries

Comparing treatment and comparison group characteristics at baseline confirmed that the two groups were well balanced (see Appendix B for full matching results). Participants in the analytic sample for the COACHES program (treatment group) had a median age of 18, were split evenly across genders, and varied in their health care needs (Table 3). As noted previously, to be included in this analysis, all sample members had to have a psychiatric condition. The high use of specialty services reflects participants' elevated need for mental health services, with 40,000 specialty visits per 1,000 beneficiaries per year. This suggests that participants might have enrolled at a time when they had a high need for services. Participants also had a range of physical health conditions, most commonly a skeletal condition (13 percent), pulmonary condition (12 percent), or cardiovascular disease (11 percent). Study sample members were at high risk for needing expensive services. Their average Chronic Illness and Disability Payment System (CDPS) score indicates that sample members should have expenditures four times the average for Medicaid, and 17 percent were hospitalized at least once in year before enrolling.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Measure	Treatment (N = 299)	Comparison (N = 573)
Demographics		
Age at enrollment, years	18	18
Male, %	50	50
Health conditions, %		
Central nervous system	5.4	5.0
Cardiovascular	11	10
Developmental disability	1.0	3.0
Gastrointestinal	5.7	7.2
Metabolic	1.3	1.6
Psychiatric	100	100
Pulmonary	12	14
Renal	1.3	1.6
Skeletal	13	9.2
CDPS score[a]		
Mean	3.9	4.0
25th percentile	2.7	2.2
Median	3.1	3.1
75th percentile	4.9	5.1
Service use during year before enrollment		
Any hospitalization at baseline, %	17	17
Number of outpatient ED visits (per 1,000 beneficiaries)	1,002	987
Number of inpatient stays (per 1,000 beneficiaries)	239	251
Ambulatory primary care encounters (per 1,000 beneficiaries)	4,389	4,078
Ambulatory specialty care encounters (per 1,000 beneficiaries)	45,257	42,573
Awardee-specific measures during year before enrollment[b]		
Any claim for long-acting birth control among females, %	39	43
Any claim for non long-acting birth control among females, %	60	56

Sources: Mathematica's analysis of information from awardee's finder file and Medicaid managed care encounter and enrollment data from February 1, 2013, through August 31, 2018, as of January 2020.

Notes: The analysis defined the baseline period as the 365 days before and including each beneficiary's enrollment date. It defined the enrollment date as the date on which the participant signed a consent form to participate in the research study. It also measured all beneficiary characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid. In addition to the number of months enrolled in FFS Medicaid, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

ED visit measures include observation stays.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending. The analysis calculated CDPS scores by using the most recently available algorithms developed by the Centers for Medicare & Medicaid Services.

[b] The analysis measured awardee-specific baseline measures only among female beneficiaries.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; HCC = hierarchical condition category.

Analytic approach

The regression estimates for continuous measures (for example hospitalizations per 1,000 participants) relied on a difference-in-differences model. This model measures program effects as the change in outcomes among study participants before versus after enrollment relative to the change in outcomes among a comparison group.

The primary outcomes were the hospitalization and emergency department (ED) visit rates per 1,000 beneficiaries, and ambulatory primary care visits and specialty care visits per 1,000 beneficiaries. Secondary outcomes related to the intervention included birth control use among female beneficiaries who were not on long-acting birth control at baseline.[2] Expenditures were not analyzed because most participants were in comprehensive managed care and plans' expenditures on services provided were not available in claims data. Appendix A describes the statistical models used to examine outcomes.

The analysis defined the pre-enrollment period as the year before each participant's enrollment date. It defined the enrollment date as the date the participant signed a program consent form. A pseudo-enrollment date was randomly assigned to each comparison beneficiary that mirrored the distribution of enrollment dates among the treatment group by age when the intervention began.

IMPACT RESULTS

The impact estimates do not provide evidence that the COACHES program had a sustained effect on most health care outcomes (Table 4). Both primary care and specialty visits increased more for participants than for the comparison group during the first 6 months of the program (by an estimated 30 and 53 percent, respectively), possibly due to COACHES improving beneficiaries' access to services while they were in the program. However, only the specialty visit finding remained sizeable and statistically significant during the full first year after enrollment; it is possible that the treatment group participants enrolled in the program when they had a high need for specialty services. Relative to the comparison group, ED visits fell more for participants over the 7 to 12 months after enrollment, but the estimated differences were not statistically significant over the first full year of enrollment. Also, although the direction of the estimated effects were favorable (but not statistically significant) for both hospitalizations and ED visits over the first year of enrollment, the estimates became unfavorable (and not statistically significant) by the second year of enrollment (see Appendix C). Descriptive analyses showed that among females who did not have any Medicaid claims for long-acting birth control at baseline, there was a 10 percentage point increase in the having any claim for birth control during the first six months after enrollment (not shown). This increase was consistent with the program's pregnancy prevention goal. (Given the small sample size of females without birth

[2] Women do not need to get long-acting contraceptives regularly, so women who were on long-acting contraceptives at baseline would not have to get them again at follow-up. Regressions were not estimated for the birth control outcome measures due to the small sample of women without a claim for long-acting birth control at baseline.

control at baseline, regression analyses on the birth control measure was not conducted). Appendix C presents the full set of regression results. Appendix D shows the results from the Bayesian analysis.

Table 4. Estimated impact of the Amerigroup COACHES intervention on select use measures during the first year after enrollment

	All beneficiaries		
	Impact estimate	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries[b]			
Baseline year			
Months 1–6	-69	-30%	0.39
Months 7–12	-3.1	-1.9%	0.57
Year 1	-39	-19%	0.53
ED or observation visits, per 1,000 beneficiaries[b]			
Baseline			
Months 1–6	195	20%	0.37
Months 7–12	-598**	-33%	0.04
Year 1	-174	-13%	0.39
Ambulatory primary care visits, per 1,000 beneficiaries			
Baseline year			
Months 1–6	952*	30%	0.06
Months 7–12	165	6.8%	0.72
Year 1	549	20%	0.20
Specialty care visits, per 1,000 beneficiaries			
Baseline year			
Months 1–6	13,450***	53%	< 0.01
Months 7–12	6,472*	40%	0.09
Year 1	9,803***	47%	< 0.01
Sample sizes			
Number of beneficiaries			
Service use regressions	Treatment 299	Comparison 573	

Sources: Mathematica's analysis of information from the awardee's program encounter data and Medicaid claims and enrollment data from February 1, 2013, through August 31, 2018, as of January 2020.

Note: Regression estimates for changes in number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The follow-up periods are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the regression outcome estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

COACHES = Coaching and Comprehensive Health Supports; ED = emergency department.

Given how Amerigroup implemented the COACHES program, the lack of a strong association between the program and participants' use of health care services is not surprising. Participants spent an average of only 3 months in the program, considerably less time than the 12 to 18 months that the awardee anticipated when it designed the program; thus, participants might not have been enrolled long enough to fully benefit from it. Program impacts might also have been impeded by the challenges the awardee experienced engaging community providers, which limited staff's ability to coordinate care across health and social service systems. Finally, the Medicaid claims data could not measure many of the outcomes the program sought to affect (including higher educational attainment and increased employment, improved connections to peer and adult social supports, and improved life skills). Although most participants engaged in the program, they focused on goals related to outcomes that this evaluation could not measure.

Main findings from impact evaluation

- The COACHES program had no discernible impact on ED visits or hospitalizations when measured over the first full year of enrollment.
- The program was associated with a large and persistent estimated increase specialty care visits, possibly driven by a high need for specialty care at enrollment.
- The absence of an estimated impact on ED visits and hospitalizations might be due to the shorter-then-intended period of enrollment, and the lack of participants' focus on health-related goals.

CONCLUSION

The COACHES program served more than 800 youth in foster care. These youth partnered with coaches who educated them about the health and social services systems and helped them achieve their goals. This analysis suggests that the program might have helped connect beneficiaries to some health care services, as participants' receipt of birth control and rate of primary care and specialty visits were high relative to the comparison group while they were enrolled in the program. However, the increase in physician visits did not translate into longer-term estimated differences in ED use and hospitalizations. The direction of the estimates for hospitalizations and ED visits was favorable (but not statistically significant) during the first full year of the program, but by Year 2, the estimated differences became unfavorable. The program was youth directed, and youth mostly focused on education and employment goals with their coaches. Because youth participated in the program for considerably less time than anticipated and focused on nonhealth related goals, the lack of an association between the COACHES program and sizeable reductions in the use of Medicaid services is not surprising.

Limitations of evaluation

The analysis has several limitations. First, the participation rate was too low to include all eligible youth in the treatment group, which meant that the treatment group included only

participants. This meant there was likely selection bias due to participants self-referring to the program, though it is unclear how that bias might affect the estimates. Self-referring beneficiaries might be more motivated and able than others to manage their own health effectively, making the estimates more favorable. Or they might have more severe health problems not measurable in claims that made them want to enroll in the program, making estimates less favorable. Furthermore, the timing of when an eligible youth chose to enroll might be related to his or her needs at the time, which claims data could not replicate and which could help explain the erratic pattern of estimated program effects over time. Second, the impact analysis excluded COACHES participants who did not have a signed HIPAA consent form to participate in the research study, did not have mental health concerns documented in the Medicaid data, had residences outside the 34 counties where the awardee implemented the program, did not have sufficient baseline or follow-up data, or were outside the awardee's stated eligible ages for enrollment (17 to 20 years). It is unknown how including these enrollees would have affected the results. These groups accounted for 65 percent of total participants, and their exclusion limits the generalizability of the evaluation to all enrollees in the COACHES program. Third, the treatment group included youth who resided in the Atlanta metropolitan area (and several rural counties); there was not a similarly sized urban area in Georgia from which to draw the comparison group. Thus, differences in outcomes due to program effects could be confounded with differences in outcomes between urban and rural areas. Finally, the final sample sizes were likely too small to detect significant changes in most measures.

PROGRAM SUSTAINABILITY

By the end of its award in February 2018, Amerigroup had fully transitioned the COACHES program to its implementation partner, Families First. Families First secured a $2 million contract from the Georgia Department of Social Services to sustain the COACHES program for about one year beyond the cooperative agreement. During this period, Families First expanded the geographic scope of the program to include additional counties and lowered the age range for the program to include youth ages 15 and 16, in addition to those ages 17 to 20.

However, Families First did not receive renewed state funding for the second year following the cooperative agreement. Starting then, Families First relied on private donations to run the program, which resulted in a decrease in the program's operating budget. As a result, Families First reduced the program's geographic reach to about one-third of the original number of counties and narrowed its focus to helping youth achieve education and employment goals only (not health-related goals).

Amerigroup initially had plans to develop a value-based payment model for Medicaid that tied to performance on designated process and outcome measures. However, because Families First had received a state contract to operate the COACHES program, Amerigroup did not further develop or pursue a payment model.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays rely on a difference-in-differences approach with beneficiary fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary birth control outcomes is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics, and uses only female sample members with no claims for birth control during the year before enrollment. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of outcomes.

The impact analysis included only 299 of the 860 total Coaching and Comprehensive Health Supports (COACHES participants (35 percent) (Table A.1). The study sample included participants only if they (1) had signed a consent form to participate in the research study during a face-to-face encounter with a coach from February 2015 to August 2017 (298 excluded participants failed to meet this requirement); (2) had sufficient data available for analysis (29 failed to meet this requirement); and (3) met the COACHES eligibility criteria based on analysis of Medicaid data (age 17 to 20 years, at least 12 months in foster care, and a behavioral health diagnosis in claims);[3] the study excluded 234 participants for not meeting one or more of these eligibility criteria.

Table A.1. Identifying the final sample for analysis for Amerigroup

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total number of reported participants from awardee		**860**
Not included in finder file (primarily due to lack of HIPAA consent)	298	562
Not found in Medicaid data or did not meet the standard claims-based criteria[a]	29	533
Did not meet program eligibility requirements:		
Not younger than 21 years	61	472
Not enrolled in foster care in their COACHES program enrollment month or not continuously enrolled in foster care in the 12-month baseline period	145	327
Did not have psychiatric or substance abuse diagnosis in baseline period	28	299
Final analytic sample		**299**

Sources: Mathematica's analysis of information from the awardee's enrollment database from February 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data from February 1, 2013, through August 31, 2018.

[a] The analysis sample excluded participants if they met one or more of the following exclusion criteria: (1) not enrolled in Medicaid for at least 90 eligible days in the baseline year; (2) dually eligible on their enrollment date (eligible for both Medicare and Medicaid) because, without Medicare claims, their outcomes could not be accurately measured;

[3] Although the analysis excluded 28 participants who did not have mental health or substance use claims in the baseline period, these participants might have had mental health concerns otherwise documented, for example in foster care case files.

Table A.1 (*continued*)

(3) were eligible for only restricted Medicaid benefits; (4) had some type of third-party coverage; (5) were enrolled in S-CHIP; or (6) died within 30 days of enrollment.

CHIP = Children's Health Insurance Program; COACHES = Coaching and Comprehensive Health Supports; HIPAA = Health Insurance Portability and Accountability Act of 1996; S-CHIP = State CHIP.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Tables B.1 shows the variables used for matching and displays the weighted means of baseline characteristics for the 299 treatment beneficiaries and the 573 matched comparison beneficiaries used in the Medicaid impact analysis. The table show the means, difference in means, percentage difference, and standardized difference for each variable, which the study calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables include demographic characteristics (age, gender, and race); health status (as measured by the Chronic Illness and Disability Payment System score), chronic condition categories, and service use before enrollment. The analysis measured variables over various specified intervals within the 12 months before enrollment in the intervention. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

Table B.1 shows the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence of the statistical significance of the difference in the means. The equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the study also performed an omnibus test in which the null hypothesis is that the treatment and matched comparison groups are balanced across all linear combinations of the covariates. The results assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes. Overall, the comparisons suggest that the two groups match well on all of the measured variables.

Table B.1. Baseline characteristics of treatment and matched comparison groups for Amerigroup

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	18 (0.06)	18 (0.04)	-0.09 (0.08)	< +/-1	-0.10	0.23	0.03
Male, %	49 (2.9)	50 (2.1)	-0.92 (4.0)	-1.9	-0.02	0.82	< 0.01
Birth control method, %							
Has long acting birth control	19 (2.3)	21 (1.5)	-2.0 (3.2)	-10	-0.05	0.53	< 0.01
Has nonlong-acting birth control	30 (2.6)	28 (1.9)	1.6 (3.7)	5.4	0.03	0.67	< 0.01
Health status and diagnoses							
CDPS score	3.9 (0.12)	4.0 (0.10)	-0.06 (0.18)	-1.4	-0.03	0.75	< 0.01
AIDS or other infectious disease, %	2.0 (0.81)	1.7 (0.46)	0.32 (1.1)	16	0.02	0.77	< 0.01
Cardiovascular disease, %	11 (1.8)	10 (1.0)	0.70 (2.5)	6.5	0.02	0.78	< 0.01
Central nervous system condition, %	5.4 (1.3)	5.0 (0.91)	0.36 (1.8)	6.8	0.02	0.84	< 0.01
Cerebrovascular condition, %	1.0 (0.58)	0.07 (0.18)	0.94 (0.60)	93	0.09	0.12	< 0.01
Developmental disability, %	1.0 (0.58)	3.0 (0.67)	-2.0 (1.2)	-200	-0.20	0.09	0.34
Diabetes, %	1.7 (0.74)	1.1 (0.46)	0.52 (0.99)	31	0.04	0.60	< 0.01
Eye condition, %	0.33 (0.33)	0.27 (0.35)	0.07 (0.46)	20	0.01	0.89	< 0.01
Gastrointestinal condition, %	5.7 (1.3)	7.2 (1.1)	-1.5 (2.0)	-26	-0.06	0.44	0.01
Genital condition, %	3.0 (0.99)	2.1 (0.46)	0.94 (1.3)	31	0.05	0.47	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Hematological condition, %	2.0 (0.81)	2.2 (0.43)	-0.20 (1.2)	-10.0	-0.01	0.86	< 0.01
Metabolic condition, %	1.3 (0.67)	1.6 (0.60)	-0.28 (1.0)	-21	-0.02	0.78	< 0.01
Pulmonary condition, %	12 (1.9)	14 (1.4)	-2.6 (2.8)	-23	-0.08	0.35	0.03
Psychiatric condition, %	100	100	0	0	0	n.a.	n.a.
Renal condition, %	1.3 (0.67)	1.6 (0.60)	-0.28 (1.0)	-21	-0.02	0.78	< 0.01
Skeletal condition, %	13 (2.0)	9.2 (1.1)	4.2 (2.6)	31	0.12	0.11	0.05
Skin condition, %	7.7 (1.5)	8.6 (0.91)	-0.91 (2.2)	-12	-0.03	0.68	< 0.01
Substance abuse, %	4.7 (1.2)	3.6 (0.69)	1.1 (1.6)	23	0.05	0.50	< 0.01
Service use							
Total hospitalizations	238 (37)	250 (27)	-13 (61)	-5.3	-0.02	0.84	< 0.01
Total hospitalizations, 3 months before enrollment	255 (66)	380 (53)	-125 (111)	-49	-0.11	0.26	0.08
Primary care visits, ambulatory setting	4,389 (282)	4,081 (173)	308 (377)	7.0	0.06	0.42	< 0.01
Primary care visits, ambulatory setting, 3 months before enrollment	4,977 (346)	4,455 (225)	522 (496)	10	0.09	0.29	0.02
Specialist visits, any setting	46,912 (1,906)	44,540 (1,124)	2,371 (2,909)	5.1	0.07	0.42	0.02
Specialist visits, ambulatory setting, 3 months before enrollment	47,166 (2,435)	40,411 (1,161)	6,755 (3,387)	14	0.16	0.05	0.13
Outpatient ED visits and observation stays	1,009 (90)	982 (59)	27 (137)	2.7	0.02	0.84	< 0.01
Outpatient ED visits and observation stays, 3 months before enrollment	1,100 (144)	1,328 (113)	-228 (239)	-21	-0.09	0.34	0.05

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Propensity score	0.48 (0.01)	0.44 (0.01)	0.04 (0.02)	8.2	0.17	0.04	0.18
Number of beneficiaries	299	573					
Omnibus test				Chi-squared statistic 182.57	Degrees of freedom 30.00	*p*-value 0.00	

Sources: Mathematica's analysis of information from the awardee's program encounter data and Medicaid claims and enrollment data from February 1, 2013, through August 30, 2017. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; SE = standard error.

n.a. = not applicable: The matching did not include this variable because all treatment and comparison group beneficiaries had to have a psychiatric condition to be in the sample.

Appendix C

Detailed results from impact estimates

This page has been left blank for double-sided copying.

Table C.1 displays the results from the impact analysis, showing estimates for the study population measured over the 12-month baseline period and two-year follow-up period. The regression estimates for continuous measures (for example hospitalizations per 1,000 beneficiaries) rely on a difference-in-differences model. This model measures program effects as the change in outcomes among study beneficiaries before versus after enrollment relative to the change in outcomes among a comparison group.

The difference-in-differences models were estimated for number of services used (per 1,000 beneficiaries, annualized). The estimated percentage change in outcome is the estimated effect divided by a counterfactual value defined as the treatment group mean minus the regression estimate. One, two, or three asterisks indicate estimates that differ statistically from zero at the 0.10, 0.05, and 0.01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the Amerigroup intervention on select use measures for all beneficiaries

	All beneficiaries				
	Treatment group mean	Comparison group mean	Estimated impact (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries					
Baseline year	239	251			
Months 1–6	163	244	-69 (81)	-30%	0.39
Months 7–12	163	178	-3.1 (75)	-1.9%	0.97
Year 1	164	214	-39 (61)	-19%	0.53
Months 13–18	238	155	95 (100)	80%	0.34
Months 19–24	284	250	46 (132)	25%	0.73
Year 2	254	186	79 (99)	56%	0.42
ED or observation visits, per 1,000 beneficiaries					
Baseline year	1,002	987			
Months 1–6	1,211	1,001	195 (217)	20%	0.37
Months 7–12	1,238	1,822	-598** (294)	-33%	0.04
Year 1	1,225	1,383	-174 (204)	-13%	0.39
Months 13–18	1,679	1,340	324 (275)	25%	0.24
Months 19–24	1,758	1,757	-14 (416)	< 1%	0.97
Year 2	1,706	1,465	226 (265)	16%	0.39
Ambulatory primary care visits, per 1,000 beneficiaries					
Baseline year	4,389	4,078			
Months 1–6	4,066	2,802	952* (500)	30%	0.06
Months 7–12	2,548	2,071	165 (469)	6.8%	0.72
Year 1	3,323	2,462	549 (430)	20%	0.20
Months 13–18	1,689	1,658	-281 (545)	-13%	0.61
Months 19–24	1,078	1,729	-962 (632)	-38%	0.13
Year 2	1,479	1,695	-527 (545)	-23%	0.33

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Estimated impact (SE)	Percentage impact[a]	*p*-value
Specialty care visits, per 1,000 beneficiaries					
Baseline year	45,257	42,573			
Months 1–6	38,704	22,570	13,450*** (3,525)	53%	< 0.01
Months 7–12	22,108	12,952	6,472* (3,826)	40%	0.09
Year 1	30,581	18,093	9,803*** (3,450)	47%	< 0.01
Months 13–18	12,930	8,401	1,845 (4,222)	15%	0.66
Months 19–24	9,331	10,048	-3,401 (4,337)	-23%	0.43
Year 2	11,725	9,068	-27 (4,079)	< 1%	0.99
Sample sizes					
Baseline year	299	570			
Months 1–6	299	570			
Months 7–12	293	520			
Year 1	299	570			
Months 13–18	255	417			
Months 19–24	160	228			
Year 2	259	422			
Cumulative	299	570			

Sources: Mathematica's analysis of information from the awardee's finder file through December 31, 2017, and Medicare claims and enrollment data through June 30, 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Note: The estimated impact for number of visits or stays is based on a difference-in-differences approach and shows the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The estimated impact for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiaries' characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; SE = standard error.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for Amerigroup were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to Amerigroup. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on two core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for two core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for Amerigroup led to a Bayesian estimate of the program's impact on hospital admissions of -9 percent (an estimated reduction of 18 hospitalizations per 1,000 beneficiaries) in the first year.

Table D.1. Comparison of frequentist and Bayesian impact estimates for Amerigroup in the first year after enrollment

	Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Hospital admissions	-39 (-158, 81)	-18 (-40, 4.3)	-7%	-19%	-9%
ED visits	-174 (-574, 226)	-129 (-281, 25)	-7%	-13%	-9%

Source: Mathematica's analysis of information from the awardee's program encounter data and Medicaid claims and enrollment data from February 1, 2013, through August 31, 2018, as of January 2020. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

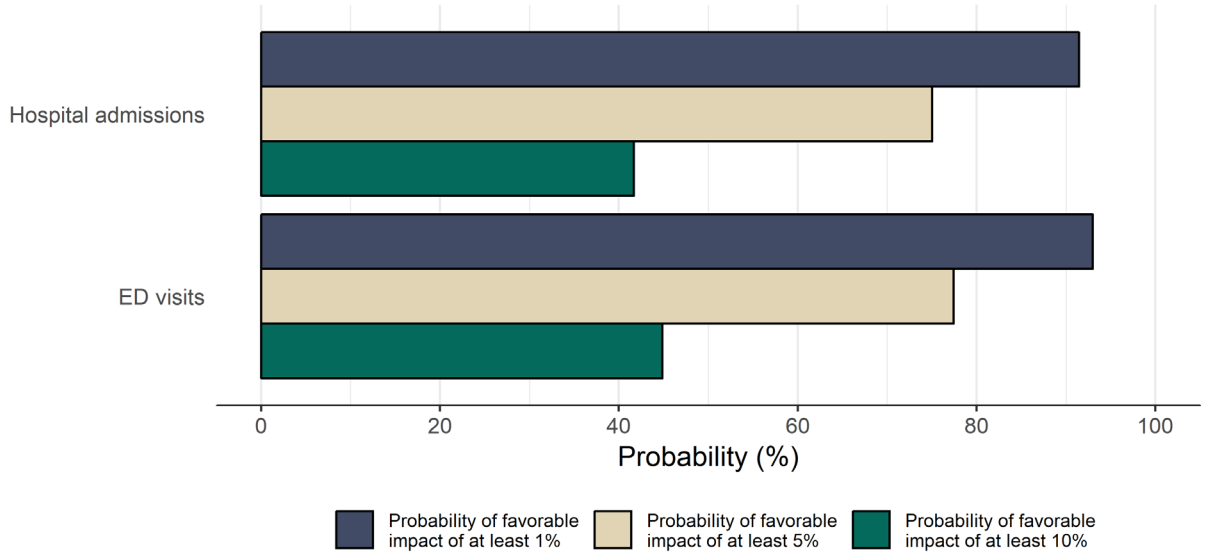
ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results relied on a small sample and are therefore imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in hospital admissions. Figure D.1 shows the probability that Amerigroup achieved favorable impacts in the first year on two core outcomes at three different thresholds: (1) a

favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the Amerigroup program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from the awardee's program encounter data and Medicaid claims and enrollment data from February 1, 2013, through August 31, 2018, as of January 2020. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a strong probability—in the range of 90 percent—that Amerigroup had a favorable impact of 1 percent or more on hospital admissions and emergency department visits, and a similarly strong probability – in the range of 70 percent – that Amerigroup had a favorable impact of 5 percent or more on these outcomes. These probabilities suggest promise but, similar to the frequentist findings, are not large enough to indicate a substantial impact.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Association of American Medical Colleges

September 2020

Brant Morefield, Vladislav Slanchev, Dana Jean-Baptiste, Ella Douglas-Durham, Bob Schmitz, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

ASSOCIATION OF AMERICAN MEDICAL COLLEGES

The Association of American Medical Colleges (AAMC) received a cooperative agreement under Round 2 of the Health Care Innovation Awards to implement the Coordinating Optimal Referral Experience (CORE) program. The goal of the CORE program was to reduce wait times for specialty care appointments and increase the effectiveness of referral processes by improving communication and care coordination between primary care physicians (PCPs) and specialists across 18 targeted specialties. The program aimed to reach all primary care clinic patients served by the five academic medical centers (AMCs) participating in the program. Table 1 summarizes the key characteristics of the program.

Important issues for understanding the evaluation

- The CORE program aimed to reduce long wait times for specialty appointments and increase the effectiveness of referral processes by improving communication and care coordination between PCPs and specialists in the 18 included specialties.
- The program embedded two templates, eConsult and eReferral, in the EMR systems of all participating AMCs.
- The impact analysis was limited to Medicare FFS beneficiaries who were potential participants and for whom claims data were available to measure the program effects on core outcomes. The awardee expected improved care coordination to reduce preventable hospitalizations, ED visits, and costs.
- In the impact analysis, a primary care visit to any participating or comparison AMC triggered a 91-day observation episode. All outcomes were measured during the 91-day episode. After the 91 days elapsed, another primary care visit could initiate a subsequent episode for the same beneficiary.

The CORE program used two templates—the eConsult and the eReferral—to support the program goals. The *eConsult* template aimed to bring guidance and clarification from a range of specialty areas to help PCPs manage a patient's care. The *eReferral* template sought to facilitate and streamline the specialty referral process by guiding the PCP on the types of information that a specialist would need beforehand. The program embedded both templates in the electronic medical record (EMR) systems at all primary care practices and community-based clinics affiliated with the AMCs. Staff tailored the templates to support use across 18 medical specialties included in the CORE program. PCPs accessed the templates through the EMR system, and specialist physicians also affiliated with the AMC later completed them. The awardee hypothesized that use of the two templates would fill gaps in communication and coordination between the primary care and specialty care providers and facilitate consultation between PCPs and specialists in treating patients. This, in turn, would help reduce unnecessary subspecialty referrals and visits, yield more efficient use of specialist care, improve access to specialists, reduce unnecessary emergency department (ED) visits and hospitalizations, and lower total expenditures for payers.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	AAMC implemented the CORE program to enhance care delivery at the primary care–specialty care interface by giving PCPs decision support templates designed to help them seek guidance about patients' treatment, to guide the PCP on the types of information that specialists needed before the referral, and to help the specialists assess whether the referral was appropriate.
Major innovation	• Formalized communication between PCPs and specialists • Leveraged the EMR system to facilitate communication and coordination • Pulled patients' data from the EMR to support the referral process • Expanded access to specialist feedback
Program components	• eConsult templates • eReferral templates • Health IT
Target population	The primary focus of the CORE program was to use PCPs employed by the AMCs at both AMC-based and community-based clinics. The target population consisted of patients ages 18 and older of any payer status who visited the primary care practice sites.
Participating providers	The program sought to reach PCPs and specialists at five participating AMCs that comprise both AMC-based and community-based clinics. The five participating AMCs included Dartmouth-Hitchcock; the University of California, San Diego; the University of Iowa; the University of Virginia; and the University of Wisconsin.
Total enrollment	The awardee enrolled 128,721 unique indirect participants.
Level of engagement	According to the awardee's self-reported monitoring data, the eConsult rate increased across participating AMCs from about 4 per 10,000 patients in September 2014 to nearly 14 per 10,000 patients in May 2017. In total, 88 percent of respondents to the clinician survey said the program increased collaboration between PCPs and specialists.
Theory of change or theory of action	AAMC hypothesized that combining improved coordination and communication between PCPs and specialists with the eConsult interface would lead to a reduction in unnecessary referrals and visits, more efficient use of specialist care, and improved access to specialists. This, in turn, would help reduce unnecessary ED visits and hospitalizations, and lower total expenditures for payers.
Award amount	$7,125,770
Effective launch date	January 1, 2015
Program settings	Primary care practices, hospitals, and AMCs
Market area	Rural, urban, and suburban
Target outcomes	• Increased patient satisfaction • Decreased ED visits and hospitalizations • Decreased total cost of care • Decreased number of referrals • Increased quality of eConsults • Decreased cost for diagnostic testing and imaging • Increased eConsult use • Increased access to specialty care • Decreased out-of-pocket costs to patients (estimated)

Table 1 (*continued*)

Program characteristics	Description
Payment model	New FFS payment
Sustainability plans	The award enabled AMCs to demonstrate the program's value to providers and patients alike. As a result, all five participating AMCs reported plans to use internal resources to sustain program components to varied degrees and through varied means. The awardee expanded the program to other member AMCs, affiliated providers, and patient populations.

AAMC = Association of American Medical Colleges; AMC = academic medical center; CORE = Coordinating Optimal Referral Experience; ED = emergency department; EMR = electronic medical record; FFS = fee-for-service; IT = information technology; PCP = primary care provider.

The impact analysis was limited to a treatment group of 541,472 Medicare fee-for-service (FFS) observations (defined as non-unique beneficiary episodes) identified following a primary care visit to a participating AMC clinic and a comparison group of 979,532 observations reflecting beneficiaries with similar characteristics who received services at 14 comparison AMCs. The analyses compared differences in treatment and comparison group outcomes before program implementation (March 2013 to February 2015) with the differences in outcomes after full implementation of the award (March 2016 to August 2017). Table 2 summarizes the key features of the impact evaluation. Appendix A describes the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The impact estimate relied on a cross-sectional, difference-in-differences model that compared changes in outcomes among episodes at participating AMCs before versus after the practice joined the program with changes in episodes over the same period at comparison AMCs.
Intervention group for evaluation	The impact analysis relied on all episodes initiated by Medicare FFS beneficiaries who visited a participating or comparison AMC for primary care before (March 2013 to February 2015) or 12 months or more after the award start date (March 2016 to August 2017), when templates for all 18 specialties were in use. The analysis defined episodes as 91-day observation periods following a primary care visit to any participating or comparison AMC. After the 91 days elapsed, another primary care visit could initiate a subsequent episode for a beneficiary. The treatment population included 231,143 episodes before award and 310,329 episodes after implementation. The comparison population included 411,132 episodes before award and 568,400 episodes after implementation.
Comparison group	The comparison population included 979,532 episodes initiated by a primary care visit at 14 comparison AMCs (411,132 during a pre-implementation period and 568,400 during a post-implementation period). Comparison AMCs were selected as the nearest neighbors from the same state for each of the five awardee AMCs, based on the number of PCPs and the number of physicians practicing in the specialties for which the templates were developed.
Limitations	Pre-post differences in adjusted outcomes between treatment and comparison AMCs in the absence of the intervention would have biased the impact estimates. In addition, estimating impacts over a broader population of all eligible patients could lead to an under-estimation of program effects.

AMC = academic medical center; FFS = fee for service; PCP = primary care physician.

PROGRAM DESIGN AND ADAPTATION

The CORE program comprised three components: (1) eConsult templates, (2) eReferral templates, and (3) health information technology (health IT).[1]

eConsult templates

The eConsult template represented an electronic information exchange initiated by a PCP who sought guidance from a specialist. Designed for individual specialties, the templates aimed to fill gaps in communication and coordination between the primary care and specialty care providers and facilitate consultation between the two across the course of treating patients. PCPs electing to use the eConsult template could pose patient-specific clinical questions regarding the patient's condition or symptoms to a specialist that would otherwise have required the PCP to refer the patient to a specialist.

eReferral templates

The eReferral template conveyed pre-specialist consultation guidance at the point of referral and supported the PCPs in understanding whether a referral was necessary. Tailored to different specialties, specialists used the eReferral template at the point of referral to help assess whether the referral was appropriate and to guide the PCP on the types of information that specialists would need before the referral visit. Information requested in the eReferral template included the basic clinical history for the patient's condition along with confirmation of whether the PCP completed certain key diagnostic and laboratory tests.

Health IT

Health IT was an important component for the program, as the eConsult and eReferral templates were embedded within the EMR systems of participating AMCs. The CORE program's innovation stemmed from the way in which it formalized and streamlined communication, pulled patients' diagnostics and medical history data, and expanded access to feedback from a specialist—all through the AMCs' EMR systems.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee exceeded its three-year enrollment target without making any major changes in its eligibility criteria, enrollment process, or the CORE program during the cooperative agreement. The awardee implemented the CORE program gradually and did not report any notable delays in the implementation process. Using a phased approach to launch the templates was intentional, enabling sites to refine and adapt the templates to local needs. The awardee addressed several

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

initial difficulties coordinating the use of the templates across the AMCs' EMR systems, but allowing for the customizations enabled the sites to implement the templates as intended. By January 2015, six months after the start of the award, each of the five AMCs had implemented the eConsult and eReferral templates in at least two specialty areas.

Implications of program implementation for detecting impacts

- Because the program implemented the eConsult and eReferral templates in phases over the first year of the award, this analysis presents impact estimates for observations made 12 or more months after the award start date.
- Because the University of California system implemented a similar design based on the program developed at the University of California at San Francisco, one participant (the University of California at San Diego) began implementing the eConsult and eReferral templates earlier than other awardee AMCs.
- The eConsult and eReferral templates were available for use during all primary care visits, though they might not have been used. In addition, physicians might have adapted their referral behavior after using the template, without directly applying the template thereafter. As a result, the impact study measured effects over all primary care visits regardless of template use.

Even with a relatively smooth implementation, several AMCs faced early challenges getting PCPs at their community-based clinics to use the eConsult and eReferral templates. The AMCs implemented various strategies to encourage template use, such as granting community clinics video access to provider comanagement conferences and increasing educational outreach to the clinic providers. The awardee also encountered initial resistance to using the eReferral templates from at least some clinicians across all AMCs because of the length and complexity of the tool. In response, AMCs began requiring all providers to use the eReferral templates so that using templates became the only way they could refer patients to specialists.

ESTIMATING PROGRAM IMPACTS

Enrolling participants

Because the CORE program did not provide services directly to patients, participant enrollment reflected a population of indirect participants passively enrolled in the program. That is, the program trained clinicians who treated the target population and considered all patients served by the PCPs at participating AMCs as indirect participants. This is because all patients were eligible to receive an eConsult or eReferral, and those who did not get one could have benefitted from the model indirectly as PCPs learned to more effectively make appropriate specialist care referrals from earlier use of the template. Indirect program participants included all patients older than 17, regardless of payer status, who visited the primary care practices.

Study sample

The impact evaluation included all 91-day primary care episodes for Medicare FFS beneficiaries who visited a participating or comparison AMC for primary care. A beneficiary's initial primary care visit to a treatment or comparison AMC triggered an episode, which included the 91 days following the visit. After 91 days, if the beneficiary had another primary care visit to a participating or comparison AMC, that could trigger a subsequent episode. Thus, Medicare patients could have more than one episode in the study.

The treatment population included 541,472 episodes initiated at the five awardee AMCs (231,143 before program implementation and 310,329 after). The comparison population included 979,532 episodes initiated by a primary care visit at 14 comparison AMCs (411,132 during a pre-implementation period and 568,400 during a post-implementation period). Comparison AMCs were selected as the nearest neighbors from the same state for each of the five awardee AMCs, based on the number of PCPs and the number of physicians practicing in the specialties for which the templates were developed. Because the University of California system implemented a similar program, the analysis drew comparison AMCs from the Pacific Census region, excluding the state of California.

Characteristics of treatment and comparison group beneficiaries

A comparison of characteristics between treatment and comparison episodes shows there was appropriate balance when using an inverse propensity score weighting approach (Table 3). The average age of treatment and comparison group beneficiaries was 71 years. Forty-two percent of the beneficiaries were male and 89 percent were White. About 19 percent of both groups were dually eligible for Medicare and Medicaid. The mean hierarchical condition category scores for treatment and comparison beneficiaries were 21 and 22 percent higher than the national average for Medicare beneficiaries. Per beneficiary per month expenditures averaged slightly more than $900 for both groups. None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level using a two-tailed test. Appendix B presents the full balance results.

Table 3. Baseline characteristics of treatment and comparison group episodes

Measure	Treatment group (N = 541,472)	Comparison group (N = 979,532)
Demographics		
Age at enrollment, years	71	71
Age group, %		
Younger than 65	17	17
65 to 74	43	43
75 to 84	27	27
85 and older	12	12
Male, %	42	42

Table 3 (*continued*)

Measure	Treatment group (N = 541,472)	Comparison group (N = 979,532)
White, %	89	89
Dual eligibility		
Medicare–Medicaid dual status, %	19	19
Health status		
HCC score[a]	1.21	1.22
Service use and expenditures during the year before enrollment		
Number of hospital admissions (per 1,000 beneficiaries)	290	292
Number of primary care visits, any setting (per 1,000 beneficiaries)	5,415	5,388
Total Medicare expenditures ($ PBPM)	910	924

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, to August 31, 2017, as of March 13, 2019.

Notes: Counts for treatment and comparison groups contain both pre- and post-intervention period episodes. The analysis defined the baseline year as the 365 days before each episode. The episode initiation was the date of a participant's primary care visit that triggered an episode. The analysis measured all beneficiaries' characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

Appendix B presents the full balance results.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CMS = Centers for Medicare & Medicaid Services; ESRD = end-stage renal disease; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact estimate relied on a cross-sectional, difference-in-differences model that compared changes in outcomes among episodes at participating AMCs before versus after the practice joined the program with changes in episodes over the same period at comparison AMCs. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. Because participating AMCs started using the eConsult and eReferral templates for varying specialties throughout the first year, the impact analysis estimated effects for beneficiaries with a primary care visit 12 months or more after the award start date to ensure that the impact estimate captured the full effects of a mature program. As mentioned, the study included all eligible 91-day episodes initiated by a primary care visit, regardless of whether the episode used an eConsult or eReferral service. The analysis used a fixed effects model to account for the correlation in outcomes between multiple episodes for the same individual.

The pre-implementation period was March 2013 to February 2015 and the post-implementation period was March 2016 to August 2017. The primary outcomes included total Medicare spending, number of hospital admissions, and number of ED visits. Because increased communication and efficiencies within the referral process could affect appropriate and

necessary referrals, secondary outcomes included number of primary care visits and whether a beneficiary had a visit with one of the specialties for which the templates were implemented. Appendix A provides additional detail on the analytic approach for estimating program impacts.

IMPACT RESULTS

The study estimated that the CORE program reduced total Medicare expenditures, inpatient expenditures, and the number of primary care visits during a three-month follow-up period by **2 percent** among treatment group episodes relative to comparison group episodes (Table 4). These results were statistically significant. The impact analysis also estimated a **2 percent** increase in the likelihood of receiving a specialist visit in the highest-volume specialties for treatment group episodes during the three-month follow-up period. But there was no significant change in the likelihood of a specialist visit among the treatment population when measured across all program specialties. Appendix C presents the full results of the impact analysis. Appendix D shows the results from the Bayesian analysis.

Table 4. Estimated impact of CORE program on selected outcomes

	Full group	Treatment estimates excluding the University of Iowa
Expenditures ($ PBPM)		
Impact ($)	-$21***	-$10
Percentage impact	-2.1%	< 1%
p-value	< 0.01	0.23
Acute inpatient expenditures ($ PBPM)		
Impact ($)	-$7*	$0
Percentage impact	-2.3%	< 1%
p-value	0.09	0.92
Number of hospitalizations, per 1,000 beneficiaries		
Impact (rate)	-4.7	-1.2
Percentage impact	-1.5%	< 1%
p-value	0.21	0.75
Number of ED visits, per 1,000 beneficiaries		
Impact (rate)	1.2	2.7
Percentage impact	< 1%	< 1%
p-value	0.84	0.67
Number of primary care visits, per 1,000 beneficiaries		
Impact (rate)	-166***	-157***
Percentage impact	-1.8%	-1.7%
p-value	< 0.01	< 0.01
Probability of having a specialist visit (cardiology, dermatology, oncology, orthopedics, or psychiatry)		
Impact (pp)	0.75***	0.96***
Percentage impact	1.9%	2.4%

Table 4 (*continued*)

	Full group	Treatment estimates excluding the University of Iowa
p-value	< 0.01	< 0.01
Probability of having a specialist visit (any award specialty)		
Impact (pp)	0.24	0.41**
Percentage impact	< 1%	< 1%
p-value	0.19	0.03
Sample size		
Treatment episodes		
Pre-implementation period	231,143	214,250
Post-implementation period	310,329	285,140
Comparison episodes		
Pre-implementation period	411,132	411,132
Post-implementation period	568,400	568,400

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, to August 31, 2017, as of March 13, 2019.

Notes: Impact estimates relied on the regression-adjusted difference between the treatment and inverse propensity weighted control group episodes. The analysis then calculated percentage impacts as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents full impact estimates. Appendix D shows the results from the Bayesian analysis.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

CORE = Coordinating Optimal Referral Experience; ED = emergency department; PBPM = per beneficiary per month; pp = percentage point.

However, sensitivity analyses show that one participating AMC (the University of Iowa) drove the statistically significant reduction in expenditure. This AMC had the highest pre- and post-implementation average expenditures among all participating and comparison AMCs. Effects estimated over the four non-Iowa participating AMCs relative to the comparison AMCs do not show significant changes in either total or inpatient expenditures during the three-month follow-up period. The sensitivity results show the same estimated reductions in primary care visits and the likelihood of a specialist visit as the basic results. Because the awardee did not expect the program to influence inpatient expenditures, and the finding is sensitive to the inclusion of one AMC, the study's robust findings reflect statistically significant changes in primary care and specialist use, but they do not support the finding of a significant reduction in expenditures.

These findings are consistent with the program's theory of action. Greater efficiency in specialist referrals and access to specialists could explain both the reduction in primary care visits and the higher likelihood of visiting a specialist. If the use of the templates increased the efficiency with which PCPs could identify and refer beneficiaries to specialists, then the program should have induced fewer repeat visits to PCPs and a greater likelihood of making appropriate specialist referrals within 91 days. Two program components in particular might have contributed to an increase in referral efficiency. First, increased decision support for PCPs might enhance their

knowledge of specialty care and reduce the need for repeat primary care visits. In fact, 86 percent of respondents to the clinician survey reported that the program increased PCPs' knowledge about issues that often require specialist input or referral. Second, by updating the referral process, eReferrals could have reduced wait times for specialty appointments.

CONCLUSION

Initial impact estimates suggested statistically significant reductions in total and inpatient Medicare spending for the participating AMCs relative to the comparison AMCs during a three-month follow-up period. But one participating AMC, which had the highest expenditures but no change in hospital stays or ED visits, drove these results. As a result, the estimated change in total expenditures was likely not a result of the program. The analysis, however, identified robust evidence of a reduced number of primary care visits for beneficiaries at participating AMCs relative to comparison AMCs and an increase in the use of the most common intervention specialties. These results could stem from increased efficiency of referrals, increased access to specialists, reduced unnecessary primary care follow-up visits, and improved coordination between specialists. However, the finding of increased specialist visits and reduced primary care visits runs counter to current literature that suggests the opposite pattern typically leads to lower total costs.

Main findings from impact evaluation

- The study estimated the number of primary care visits experienced by beneficiaries in the CORE program during the 91 days following a primary care visit at a participating AMC to drop by nearly 2 percent relative to the change among comparison group beneficiaries treated at nonparticipating AMCs.
- It estimated treatment beneficiaries to be nearly 2 percent more likely to have a specialist visit for one of the five most commonly visited specialties for which eConsults and eReferrals were available.
- Analyses showed statistically significant reductions of $21 in total Medicare spending and $7 in inpatient spending among the treatment group relative to the comparison group. The size of the estimates and their statistical significance drops markedly if one AMC (the University of Iowa) is excluded.

Limitations of evaluation

The analysis has several limitations. First, the cross-sectional difference-in-differences estimator used for impact findings assumes that pre-post differences in adjusted outcomes would be equivalent for treatment and comparison AMCs in the absence of the intervention. Failure of this assumption would lead to biased estimates of the program. Second, the impact analysis included all patients of PCPs, whether they were appropriate candidates for use of the eConsult or eReferral templates. The use of a broader population in the model might bias estimated impacts toward zero, potentially diluting the observed effect of the templates' use. Third, treatment and comparison AMCs operate in environments with numerous other payer demonstrations and alternative payment models that affect primary care, specialist use, and expenditures. Other

activities at the treatment or comparison AMCs generating changes over the observation period could confound estimated effects of the Health Care Innovation Award.

PROGRAM SUSTAINABILITY

By the end of its award in August 2018, AAMC planned to sustain the CORE program in its entirety (both eConsult and eReferral services). Each of the five health systems participating in the program worked with its leadership to embed the program into regular operations of an existing department and adjusted staffing to maintain quality oversight. Most of the AMCs used internal funding to reimburse PCPs and specialists for the program. Many of these health systems also continued to add to the types of specialties that can participate in the program, and some extended the program to PCPs outside their systems. The awardee also expanded the program to six additional AMCs and six children's hospitals.

By the end of the award period, AAMC also progressed on developing Medicare FFS codes to generate external funding to support the program. With assistance from the University of California at San Francisco and a consultant on ways to develop the codes, AAMC successfully encouraged the Center for Medicare & Medicaid Innovation and the Relative Value Scale Update Committee to propose two new billing codes for the 2019 Medicare physician fee schedule that would reimburse both the PCP and the specialist in their role in eConsults. AAMC will provide feedback on proposed guidelines during the public comment period to help prevent misuse of the codes. Further, AAMC is exploring potential payment for eConsults with five state Medicaid programs, but discussions with private payers to pilot an eConsult payment stalled.

AAMC's proposed payment model

AAMC proposed a FFS payment model, requesting the Centers for Medicare & Medicaid Services to develop new codes to bill for eConsults. The billing codes would reimburse PCPs and specialists for electronic communication with each other, such as the program's eConsult service. This was an update from previous versions of the payment model, which proposed using temporary Medicare codes to reimburse the specialist, who could then reimburse the PCP.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates rely on a cross-sectional difference-in-differences approach. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The observations in each period identify 91-day episodes initiated by a visit to a primary care physician at a treatment or comparison academic medical center physician's office. Comparison observations were weighted using inverse propensity score weights. Appendix A of Volume I of this report provides details on the general modeling strategy and the standard set of core outcomes used for this evaluation.

In addition to the standard outcomes, the study estimated impacts on the number of primary care physician visits during the 91-day episode and two indicator variables for whether the beneficiary had one or more visits from (1) the most-commonly visited specialties included in the program and (2) any of the 18 specialties for which the eConsult and eReferral templates were implemented at the five participating academic medical centers. Information about the specialties and the associated provider specialty codes for all participating AMCs is available upon request.

The number of treatment group beneficiaries included in the AAMC impact analysis exceeded the awardee's count of the total number of beneficiaries served by the program because the attribution rules used in the impact analysis, which were based on the program enrollment criteria, differed from the process of reporting by the AMCs participating in the program. In addition, the impact analysis used several methods to identify individual providers participating in the program and chose the approach that most closely aligned the reports provided by the awardee. Despite these efforts, the final list of treatment providers was not an exact match to the awardee report.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. The table displays the weighted means of baseline characteristics for the 541,472 treatment episodes and the 979,532 comparison episodes used in the impact analysis. The table shows the means, difference in means, percentage difference, and standardized difference for each variable. The analysis calculated the standardized difference as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The propensity score variables included demographic characteristics (age, gender, and race); Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category score); number of hospital admissions (12 months and 30 days before episode); number of emergency department (ED) or observation visits (12 months and 30 days before episode); number of primary care visits (12 months and 30 days before episode); and Medicare expenditures in total (12 months and 30 days before episode). The analysis used inverse propensity score weights for the comparison episodes. It measured the variables over various specified intervals within the 12 months before episode initiation. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report

The table also shows the results of the equivalency-of-means tests. The *p*-values come from a weighted two-sample t-test, which provides evidence of a statistically significant difference in the means. The equivalence test *p*-values are the greater of the two one-sided weighted t-test *p*-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the analysis performed an omnibus test in which the null hypothesis is that the treatment and weighted comparison groups are balanced across all linear combinations of the covariates. It used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

Table B.1. Baseline characteristics of treatment and matched comparison groups for AAMC

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	71 (0.02)	71 (0.01)	0.00 (0.02)	< +/-1	0.00	0.98	< 0.01
Female, %	58 (0.07)	58 (0.05)	-0.03 (0.09)	< +/-1	0.00	0.74	< 0.01
Age: younger than 65, %	17 (0.05)	17 (0.04)	-0.03 (0.07)	< +/-1	0.00	0.65	< 0.01
Age: 65 to 74, %	43 (0.07)	43 (0.05)	0.07 (0.10)	< +/-1	0.00	0.44	< 0.01
Age: 75 to 84, %	27 (0.06)	27 (0.04)	-0.05 (0.09)	< +/-1	0.00	0.61	< 0.01
Age: 85 and older, %	12 (0.04)	12 (0.03)	0.00 (0.06)	< +/-1	0.00	0.95	< 0.01
White, %	89 (0.04)	89 (0.04)	-0.07 (0.06)	< +/-1	0.00	0.24	< 0.01
Black, %	5.5 (0.03)	5.5 (0.03)	0.03 (0.04)	< +/-1	0.00	0.46	< 0.01
Hispanic, %	0.82 (0.01)	0.82 (0.01)	0.00 (0.02)	< +/-1	0.00	0.87	< 0.01
Other, %	3.0 (0.02)	3.0 (0.02)	0.03 (0.03)	< +/-1	0.00	0.43	< 0.01
Unknown, %	1.4 (0.02)	1.4 (0.01)	0.01 (0.02)	< +/-1	0.00	0.79	< 0.01
Medicare entitlement and dual eligibility status, %							
Dually eligible for Medicare and Medicaid	19 (0.05)	19 (0.04)	0.03 (0.07)	< +/-1	0.00	0.65	< 0.01
Health status and diagnoses							
HCC score[a]	1.21 (0.00)	1.22 (0.00)	0.00 (0.00)	< +/-1	0.00	0.18	< 0.01
Medicare expenditures							
Total expenditures	910 (3.0)	924 (2.2)	-14 (4.4)	-1.5	-0.01	< 0.01	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total expenditures[b]	828 (2.0)	832 (1.6)	-4.1 (2.8)	< +/-1	0.00	0.14	< 0.01
Total expenditures, 3 months before enrollment	978 (4.5)	997 (3.4)	-19 (6.8)	-2.0	-0.01	< 0.01	< 0.01
Total expenditures, 3 months before enrollment[b]	828 (2.8)	836 (2.2)	-7.9 (3.7)	< +/-1	0.00	0.03	< 0.01
Total expenditures, day of enrollment	5,154 (32)	5,078 (25)	76 (47)	1.5	0.00	0.11	< 0.01
Total expenditures, day of enrollment[b]	3,939 (4.0)	3,879 (2.8)	60 (5.7)	1.5	0.02	< 0.01	< 0.01
Total expenditures, 30 days after enrollment	1,346 (6.9)	1,404 (5.4)	-58 (10)	-4.3	-0.01	< 0.01	< 0.01
Total expenditures, first quarter after enrollment	1,215 (5.0)	1,260 (4.0)	-45 (7.4)	-3.7	-0.01	< 0.01	< 0.01
Acute inpatient expenditures, day of enrollment	743 (26)	754 (21)	-11 (39)	-1.5	0.00	0.78	< 0.01
Acute inpatient expenditures, 30 days after enrollment	454 (4.7)	487 (3.8)	-33 (7.0)	-7.3	-0.01	< 0.01	< 0.01
Inpatient other expenditures, day of enrollment	12 (3.3)	15 (3.1)	-2.9 (4.8)	-25	0.00	0.55	< 0.01
Inpatient other expenditures, 30 days after enrollment	33 (1.2)	37 (1.1)	-4.0 (1.8)	-12	0.00	0.03	< 0.01
Outpatient expenditures, day of enrollment	1,766 (8.3)	1,676 (6.2)	89 (12)	5.1	0.01	< 0.01	< 0.01
Outpatient expenditures, 30 days after enrollment	355 (1.8)	353 (1.2)	1.8 (2.5)	< +/-1	0.00	0.48	< 0.01
Physician services expenditures, day of enrollment	2,447 (6.3)	2,465 (3.1)	-17 (8.0)	< +/-1	0.00	0.03	< 0.01
Physician services expenditures, 30 days after enrollment	314 (1.3)	335 (0.87)	-21 (2.1)	-6.6	-0.02	< 0.01	< 0.01
Home health expenditures, day of enrollment	30 (2.4)	35 (2.1)	-4.0 (3.6)	-13	0.00	0.27	< 0.01
Home health expenditures, 30 days after enrollment	49 (0.56)	61 (0.49)	-12 (0.86)	-25	-0.03	< 0.01	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Skilled nursing facility expenditures, day of enrollment	18 (2.8)	11 (1.9)	7.1 (3.5)	39	0.00	0.04	< 0.01
Skilled nursing facility expenditures, 30 days after enrollment	63 (1.2)	60 (0.88)	2.4 (1.6)	3.8	0.00	0.15	< 0.01
Hospice expenditures, day of enrollment	8.2 (1.5)	4.8 (0.83)	3.5 (1.9)	42	0.00	0.07	< 0.01
Hospice expenditures, 30 days after enrollment	14 (0.40)	9.1 (0.24)	4.9 (0.51)	35	0.02	< 0.01	< 0.01
Durable medical equipment expenditures, day of enrollment	42 (1.7)	45 (1.6)	-3.5 (2.7)	-8.5	0.00	0.20	< 0.01
Durable medical equipment expenditures, 30 days after enrollment	28 (0.36)	30 (0.35)	-2.9 (0.59)	-11	-0.01	< 0.01	< 0.01
Service use							
Total hospitalizations	290 (1.1)	292 (0.95)	-2.3 (1.5)	< +/-1	0.00	0.14	< 0.01
Total hospitalizations[b]	264 (0.87)	265 (0.71)	-0.53 (1.2)	< +/-1	0.00	0.66	< 0.01
Total hospitalizations, 3 months before enrollment	323 (1.8)	325 (1.5)	-2.1 (2.5)	< +/-1	0.00	0.40	< 0.01
Any hospitalization, day of enrollment, %	0.25 (0.01)	0.25 (0.01)	0.00 (0.01)	-2.0	0.00	0 .61	< 0.01
Any hospitalization, 30 days after enrollment, %	3.4 (0.02)	3.6 (0.02)	-0.21 (0.04)	-6.3	-0.01	< 0.01	< 0.01
Total hospitalizations, first quarter after enrollment	390 (2.1)	416 (1.7)	-26 (3.0)	-6.5	-0.02	< 0.01	< 0.01
Total ED or observation visits	667 (2.6)	688 (2.1)	-21 (4.6)	-3.1	-0.01	< 0.01	< 0.01
Total ED or observation visits[b]	581 (1.5)	559 (1.2)	22 (2.1)	3.9	0.02	< 0.01	< 0.01
Total ED or observation visits, 3 months before enrollment	734 (3.6)	754 (2.8)	-19 (5.7)	-2.6	-0.01	< 0.01	< 0.01
Total ED or observation visits, 3 months before enrollment[b]	638 (2.4)	620 (1.9)	19 (3.4)	2.9	0.01	< 0.01	< 0.01
Total ED or observation visits, first quarter after enrollment	732 (3.7)	732 (2.9)	0.80 (5.6)	<+/-1	0.00	0.89	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Any ED or observation visit, day of enrollment, %	0.52 (0.01)	0.51 (0.01)	0.01 (0.01)	1.0	0.00	0.71	< 0.01
Primary care visits, any setting	5,388 (7.9)	5,415 (6.8)	-27 (11)	< +/-1	0.00	0.02	< 0.01
Primary care visits, any setting, 3 months before enrollment	4,130 (11)	4,154 (9.8)	-23 (16)	< +/-1	0.00	0.14	< 0.01
Primary care visits, any setting, first quarter after enrollment	9,530 (13)	9,905 (12)	-375 (20)	-3.9	-0.04	< 0.01	< 0.01
Any outpatient ED or observation visit, 30 days after enrollment, %	5.4 (0.03)	5.2 (0.02)	0.21 (0.04)	3.9	0.01	< 0.01	< 0.01
Any outpatient visit, day of enrollment, %	53 (0.07)	53 (0.05)	0.22 (0.10)	< +/-1	0.00	0.02	< 0.01
Any outpatient visit, 30 days after enrollment, %	72 (0.06)	71 (0.05)	1.5 (0.08)	2.1	0.03	< 0.01	< 0.01
Any physician visit, day of enrollment, %	98 (0.02)	99 (0.01)	-0.86 (0.02)	< +/-1	-0.06	< 0.01	< 0.01
Any physician visit, 30 days after enrollment, %	99 (0.01)	99 (0.01)	-0.50 (0.02)	< +/-1	-0.05	< 0.01	< 0.01
Hospice use in first quarter after enrollment, %	0.71 (0.01)	0.58 (0.01)	0.13 (0.02)	18	0.02	< 0.01	< 0.01
Number of beneficiaries	541,472	979,532					
Omnibus test				Chi-squared statistic 32,231.24	Degrees of freedom 57.00	*p*-value 0.00	

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, to August 31, 2017, as of March 13, 2019.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation; *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by inverse propensity score weighting based on logistic regression. The matching weight is 1 for beneficiaries in the treatment group and $\frac{\hat{p}_i}{1-\hat{p}_i}$ for beneficiaries in the comparison group, where $\hat{p}_i$ is the estimated propensity score for beneficiary.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

AAMC = Association of American Medical Colleges; ED = emergency department; HCC = hierarchical condition category; SE = standard error.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Tables C.1, C.2, C.3, and C.4 display the results from the impact analysis. Table C.1 shows the impact estimates for all five awardees during the three-month follow-up period, measured separately for the full sample and for the maturity sample (that is, beneficiaries who enrolled more than a year after the beginning of the intervention). Table C.2 includes similar results for the four awardees without the University of Iowa. Table C.3 shows impact estimates for all five awardees during the four- to six-month follow-up period, measured separately for the full sample and for the maturity sample, and Table C.4 includes impact estimates for all five awardees separately for the full sample and the maturity sample. The study estimated models over Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. It presents results for the proportion of beneficiaries with a 30-day readmission among all discharges, as well. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the AAMC intervention on select Medicare FFS expenditures (dollars PBPM) and use measures during a three-month follow-up period

	Overall impacts					Impacts after 12 months of program operation				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)[b]										
Months 1 to 3 (pre-intervention)	1,005	1,011				1,005	1,011			
Months 1 to 3 (post-intervention)	1,019	1,046	-20*** (7.3)	-2.0%	< 0.01	1,020	1,048	-21*** (7.9)	-2.1%	< 0.01
Acute inpatient expenditures ($ PBPM)										
Months 1 to 3 (pre-intervention)	309	291				309	290			
Months 1 to 3 (post-intervention)	301	290	-7.8** (3.9)	-2.5%	0.04	298	287	-7.0* (4.2)	-2.3%	0.09
Primary care visits in any setting, per 1,000 beneficiaries										
Months 1 to 3 (pre-intervention)	9,123	8,997				9,123	8,998			
Months 1 to 3 (post-intervention)	8,959	8,978	-145*** (24)	-1.6%	< 0.01	8,931	8,971	-166*** (26)	-1.8%	< 0.01
Any visits in cardiology, dermatology, oncology, orthopedics, or psychiatry										
Months 1 to 3 (pre-intervention)	41	44				41	44			
Months 1 to 3 (post-intervention)	41	44	0.69*** (0.16)	1.7%	< 0.01	41	44	0.75*** (0.18)	1.9%	< 0.01
Any visits in any of the shared intervention specialties										
Months 1 to 3 (pre-intervention)	56	60				56	60			
Months 1 to 3 (post-intervention)	56	60	0.17 (0.17)	< 1%	0.30	56	60	0.24 (0.18)	< 1%	0.19
Hospital stays, per 1,000 beneficiaries										
Months 1 to 3 (pre-intervention)	306	331				306	331			
Months 1 to 3 (post-intervention)	299	330	-5.4 (3.5)	-1.8%	0.13	296	326	-4.7 (3.8)	-1.5%	0.21

Table C.1 (*continued*)

	Overall impacts					Impacts after 12 months of program operation				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED or observation visits, per 1,000 beneficiaries										
Months 1 to 3 (pre-intervention)	568	518				568	518			
Months 1 to 3 (post-intervention)	565	513	3.4 (5.6)	< 1%	0.55	561	510	1.2 (6.1)	< 1%	0.84
Percentage with a 30-day readmission among all discharges										
Months 1 to 3 (pre-intervention)	15	15				15	15			
Months 1 to 3 (post-intervention)	15	16	-0.37 (0.39)	-2.4%	0.34	16	15	0.08 (0.42)	< 1%	0.85
Number of index discharges for readmissions										
Months 1 to 3 (pre-intervention)	22,343	45,219				22,343	45,219			
Months 1 to 3 (post-intervention)	43,374	91,487				28,897	60,850			
Sample sizes										
Number of beneficiaries										
Months 1 to 3 (pre-intervention)	231,143	411,132				231,143	411,132			
Months 1 to 3 (post-intervention)	461,634	843,043				310,329	568,400			

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, to August 31, 2017, as of March 13, 2019.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

AAMC = Association of American Medical Colleges; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SE = standard error.

Table C.2. Estimated impact of the AAMC intervention on select Medicare FFS expenditures (dollars PBPM) and use measures during a three-month follow-up period, excluding University of Iowa AMC

	Overall impacts					Impacts after 12 months of program operation				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)[b]										
Months 1 to 3 (pre-intervention)	976	986				976	986			
Months 1 to 3 (post-intervention)	992	1,017	-14* (7.4)	-1.4%	0.05	998	1,018	-9.6 (8.0)	< 1%	0.23
Acute inpatient expenditures ($ PBPM)										
Months 1 to 3 (pre-intervention)	293	277				293	277			
Months 1 to 3 (post-intervention)	286	275	-4.4 (4.0)	-1.5%	0.27	286	271	-0.40 (4.2)	< 1%	0.92
Primary care visits in any setting, per 1,000 beneficiaries										
Months 1 to 3 (pre-intervention)	9,046	8,922				9,046	8,923			
Months 1 to 3 (post-intervention)	8,867	8,884	-141*** (24)	-1.6%	< 0.01	8,843	8,876	-157*** (26)	-1.7%	< 0.01
Any visits in cardiology, dermatology, oncology, orthopedics, or psychiatry										
Months 1 to 3 (pre-intervention)	40	44				40	44			
Months 1 to 3 (post-intervention)	40	43	0.80*** (0.17)	2.0%	< 0.01	41	43	0.96*** (0.19)	2.4%	< 0.01
Any visits in any of the shared intervention specialties										
Months 1 to 3 (pre-intervention)	56	60				56	60			
Months 1 to 3 (post-intervention)	55	59	0.27 (0.17)	< 1%	0.11	56	59	0.41** (0.19)	< 1%	0.03
Hospital stays, per 1,000 beneficiaries										
Months 1 to 3 (pre-intervention)	294	321				294	321			
Months 1 to 3 (post-intervention)	287	318	-3.9 (3.6)	-1.4%	0.27	286	314	-1.2 (3.8)	< 1%	0.75

Table C.2 (*continued*)

	Overall impacts					Impacts after 12 months of program operation				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED or observation visits, per 1,000 beneficiaries										
Months 1 to 3 (pre-intervention)	556	506				556	506			
Months 1 to 3 (post-intervention)	551	498	3.3 (5.7)	< 1%	0.56	549	496	2.7 (6.2)	< 1%	0.67
Percentage with a 30-day readmission among all discharges										
Months 1 to 3 (pre-intervention)	15	15				15	15			
Months 1 to 3 (post-intervention)	15	15	-0.23 (0.40)	-1.5%	0.57	15	15	0.28 (0.44)	1.9%	0.52
Number of index discharges for readmissions										
Months 1 to 3 (pre-intervention)	19,896	45,219				19,896	45,219			
Months 1 to 3 (post-intervention)	38,362	91,487				25,653	60,850			
Sample sizes										
Number of beneficiaries										
Months 1-3 (pre-intervention)	214,250	411,132				214,250	411,132			
Months 1-3 (post-intervention)	424,820	843,043				285,140	568,400			

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, to August 31, 2017, as of March 13, 2019.
Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.
[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.
* Significantly different from zero at the .10 level, two-tailed test.
** Significantly different from zero at the .05 level, two-tailed test.
*** Significantly different from zero at the .01 level, two-tailed test.
AAMC = Association of American Medical Colleges; AMC = academic medical center; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SE = standard error.

Table C.3. Estimated impact of the AAMC intervention on select Medicare FFS expenditures (dollars PBPM) and use measures four to six months after enrollment

	Overall impacts					Impacts after 12 months of program operation				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)[b]										
Months 4 to 6 (pre-intervention)	885	906				885	906			
Months 4 to 6 (post-intervention)	910	926	4.6 (7.4)	< 1%	0.54	914	926	8.8 (7.9)	< 1%	0.27
Acute inpatient expenditures ($ PBPM)										
Months 4 to 6 (pre-intervention)	263	247				263	246			
Months 4 to 6 (post-intervention)	262	241	5.8 (3.7)	2.3%	0.12	262	238	7.7* (4.0)	3.0%	0.05
Primary care visits in any setting, per 1,000 beneficiaries										
Months 4 to 6 (pre-intervention)	5,771	5,856				5,771	5,857			
Months 4 to 6 (post-intervention)	5,659	5,766	-23 (25)	< 1%	0.36	5,644	5,751	-22 (27)	< 1%	0.41
Any visits in cardiology, dermatology, oncology, orthopedics, or psychiatry										
Months 4 to 6 (pre-intervention)	37	41				37	41			
Months 4 to 6 (post-intervention)	37	41	0.77*** (0.16)	2.1%	< 0.01	38	41	0.83*** (0.18)	2.3%	< 0.01
Any visits in any of the shared intervention specialties										
Months 4 to 6 (pre-intervention)	51	57				51	57			
Months 4 to 6 (post-intervention)	51	57	0.53*** (0.17)	1.0%	< 0.01	51	57	0.48** (0.19)	< 1%	0.01
Hospital stays, per 1,000 beneficiaries										
Months 4 to 6 (pre-intervention)	267	283				267	283			
Months 4 to 6 (post-intervention)	263	278	1.2 (3.4)	< 1%	0.72	261	275	2.6 (3.6)	1.0%	0.47

Table C.3 (*continued*)

	Overall impacts					Impacts after 12 months of program operation				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED or observation visits, per 1,000 beneficiaries										
Months 4 to 6 (pre-intervention)	531	489				531	489			
Months 4 to 6 (post-intervention)	530	486	3.0 (5.5)	< 1%	0.59	528	485	2.1 (6.0)	< 1%	0.73
Percentage with a 30-day readmission among all discharges										
Months 4 to 6 (pre-intervention)	17	18				17	18			
Months 4 to 6 (post-intervention)	17	17	0.35 (0.45)	2.1%	0.43	17	17	0.69 (0.49)	4.2%	0.16
Number of index discharges for readmissions										
Months 4 to 6 (pre-intervention)	17,927	38,006				17,927	38,006			
Months 4 to 6 (post-intervention)	35,532	77,101				23,836	51,369			
Sample sizes										
Number of beneficiaries										
Months 4-6 (pre-intervention)	228,350	403,316				228,350	403,316			
Months 4-6 (post-intervention)	455,221	826,121				306,150	557,031			

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, to August 31, 2017, as of March 13, 2019.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

AAMC = Association of American Medical Colleges; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SE = standard error.

Table C.4. Estimated impact of the AAMC intervention on select Medicare FFS expenditures (dollars PBPM) and use measures during a three-month follow-up period, estimated separately for each participating AMC

	Overall impacts					Impacts after 12 months of program operation				
	Pre-intervention	Post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value	Pre-intervention	12+ months post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)[b]										
Months 1 to 3 (Comparison = 1)	1,011	1,046				1,011	1,046			
Months 1 to 3 (ADA = 1)	1,253	1,255	-39* (21)	-3.0%	0.06	1,253	1,263	-39* (23)	-3.0%	0.08
Months 1 to 3 (DAK = 1)	928	954	2.3 (11)	< 1%	0.84	928	965	13 (12)	< 1%	0.29
Months 1 to 3 (UIO = 1)	1,373	1,331	-98*** (25)	-6.9%	< 0.01	1,373	1,267	-163*** (27)	-7.2%	< 0.01
Months 1 to 3 (UNV = 1)	960	979	-15 (13)	-1.5%	0.23	960	987	-8.1 (14)	-1.5%	0.55
Months 1 to 3 (UWO = 1)	938	947	-21* (11)	-2.1%	0.07	938	945	-21* (12)	-2.1%	0.08
Acute inpatient expenditures ($ PBPM)										
Months 1 to 3 (Comparison = 1)	291	290				291	290			
Months 1 to 3 (ADA = 1)	376	335	-42*** (11)	-11%	< 0.01	376	332	-45*** (12)	-11%	< 0.01
Months 1 to 3 (DAK = 1)	246	256	13** (5.7)	5.2%	0.03	246	259	20*** (6.1)	5.1%	< 0.01
Months 1 to 3 (UIO=1)	514	468	-52*** (15)	-10%	< 0.01	514	429	-88*** (15)	-11%	< 0.01
Months 1 to 3 (UNV = 1)	309	291	-15** (6.9)	-4.9%	0.03	309	290	-12 (7.3)	-4.9%	0.11
Months 1 to 3 (UWO = 1)	297	294	1.5 (6.1)	< 1%	0.80	297	293	4.8 (6.5)	< 1%	0.46
Primary care visits in any setting, per 1,000 beneficiaries										
Months 1 to 3 (Comparison = 1)	8,997	8,978				8,997	8,978			
Months 1 to 3 (ADA = 1)	9,945	9,597	-250*** (64)	-2.5%	< 0.01	9,945	9,533	-309*** (69)	-2.6%	< 0.01

Table C.4 (*continued*)

	Overall impacts					Impacts after 12 months of program operation				
	Pre-intervention	Post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value	Pre-intervention	12+ months post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 1 to 3 (DAK = 1)	9,358	9,103	-135*** (41)	-1.5%	< 0.01	9,358	9,117	-102** (44)	-1.5%	0.02
Months 1 to 3 (UIO = 1)	10,100	10,021	-204*** (74)	-2.0%	< 0.01	10,100	9,924	-286*** (80)	-2.0%	< 0.01
Months 1 to 3 (UNV = 1)	8,481	8,509	-21 (38)	< 1%	0.58	8,481	8,478	-60 (41)	< 1%	0.14
Months 1 to 3 (UWO = 1)	8,874	8,669	-203*** (36)	-2.3%	< 0.01	8,874	8,629	-230*** (39)	-2.3%	< 0.01
Any visits in cardiology, dermatology, oncology, orthopedics, or psychiatry										
Months 1 to 3 (Comparison = 1)	44	44				44	44			
Months 1 to 3 (ADA = 1)	47	48	1.7*** (0.43)	3.6%	< 0.01	47	49	1.9*** (0.47)	3.6%	< 0.01
Months 1 to 3 (DAK = 1)	42	42	1.2*** (0.26)	2.9%	< 0.01	42	42	1.6*** (0.29)	2.9%	< 0.01
Months 1 to 3 (UIO = 1)	45	43	-0.76 (0.49)	-1.7%	0.12	45	42	-1.8*** (0.54)	-1.8%	< 0.01
Months 1 to 3 (UNV = 1)	41	43	2.5*** (0.29)	6.2%	< 0.01	41	43	2.7*** (0.32)	6.1%	< 0.01
Months 1 to 3 (UWO = 1)	37	35	-1.2*** (0.25)	-3.4%	< 0.01	37	34	-1.4*** (0.28)	-3.4%	< 0.01
Any visits in any of the shared intervention specialties										
Months 1 to 3 (Comparison = 1)	60	60				60	60			
Months 1 to 3 (ADA = 1)	64	66	1.9*** (0.41)	3.0%	< 0.01	64	66	2.5*** (0.45)	3.0%	< 0.01
Months 1 to 3 (DAK = 1)	56	56	0.64** (0.26)	1.1%	0.02	56	56	0.85*** (0.29)	1.1%	< 0.01
Months 1 to 3 (UIO = 1)	59	58	-1.0** (0.48)	-1.8%	0.03	59	57	-1.8*** (0.53)	-1.8%	< 0.01
Months 1 to 3 (UNV = 1)	55	57	2.0*** (0.29)	3.6%	< 0.01	55	57	2.2*** (0.32)	3.6%	< 0.01

Table C.4 (*continued*)

	Overall impacts					Impacts after 12 months of program operation				
	Pre-intervention	Post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value	Pre-intervention	12+ months post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 1 to 3 (UWO = 1)	52	50	-2.0*** (0.26)	-4.0%	< 0.01	52	50	-2.2*** (0.29)	-4.0%	< 0.01
Hospital stays, per 1,000 beneficiaries										
Months 1 to 3 (Comparison = 1)	331	330				331	330			
Months 1 to 3 (ADA = 1)	329	290	-39*** (9.2)	-12%	< 0.01	329	285	-43*** (9.7)	-12%	< 0.01
Months 1 to 3 (DAK = 1)	265	275	13** (5.4)	5.1%	0.01	265	277	19*** (5.8)	5.1%	< 0.01
Months 1 to 3 (UIO = 1)	450	431	-24** (12)	-5.4%	0.04	450	404	-47*** (13)	-5.7%	< 0.01
Months 1 to 3 (UNV = 1)	312	294	-15** (6.1)	-4.9%	0.01	312	293	-13* (6.5)	-4.9%	0.05
Months 1 to 3 (UWO = 1)	297	293	1.2 (5.5)	< 1%	0.83	297	291	3.5 (5.8)	< 1%	0.55
ED or observation visits, per 1,000 beneficiaries										
Months 1 to 3 (Comparison = 1)	518	513				518	513			
Months 1 to 3 (ADA = 1)	649	674	31** (15)	4.9%	0.04	649	665	22 (17)	4.9%	0.20
Months 1 to 3 (DAK = 1)	552	530	-14 (9.0)	-2.6%	0.12	552	526	-15 (9.7)	-2.6%	0.11
Months 1 to 3 (UIO = 1)	723	727	0.86 (18)	< 1%	0.96	723	701	-18 (19)	< 1%	0.35
Months 1 to 3 (UNV = 1)	601	564	-19* (10)	-3.2%	0.06	601	566	-16 (11)	-3.2%	0.14
Months 1 to 3 (UWO = 1)	489	518	29*** (8.4)	5.9%	< 0.01	489	515	30*** (9.1)	6.0%	< 0.01
Percentage with a 30-day readmission among all discharges										
Months 1 to 3 (Comparison = 1)	15	16				15	16			
Months 1 to 3 (ADA = 1)	17	16	-0.76 (0.98)	-4.5%	0.44	17	16	-0.15 (1.1)	-4.4%	0.89

Table C.4 (*continued*)

	Overall impacts					Impacts after 12 months of program operation				
	Pre-intervention	Post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value	Pre-intervention	12+ months post-intervention	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 1 to 3 (DAK = 1)	15	15	-0.14 (0.65)	< 1%	0.82	15	15	0.53 (0.71)	< 1%	0.45
Months 1 to 3 (UIO = 1)	16	16	-1.4 (0.95)	-7.8%	0.15	16	16	-1.4 (1.0)	-7.8%	0.16
Months 1 to 3 (UNV = 1)	17	16	-0.48 (0.72)	-2.9%	0.51	17	16	-0.38 (0.78)	-2.9%	0.63
Months 1 to 3 (UWO = 1)	14	14	0.06 (0.59)	< 1%	0.91	14	15	0.71 (0.65)	< 1%	0.27
Number of index discharges for readmissions										
Months 1 to 3 (ADA = 1)	2,482	4,316				2,482	2,828			
Months 1 to 3 (DAK = 1)	5,611	11,126				5,611	7,487			
Months 1 to 3 (UIO = 1)	2,447	5,012				2,447	3,244			
Months 1 to 3 (UNV = 1)	5,399	10,088				5,399	6,716			
Months 1 to 3 (UWO = 1)	6,404	12,832				6,404	8,622			
Sample sizes										
Number of beneficiaries										
Months 1 to 3 (ADA = 1)	24,140	48,616				24,140	32,649			
Months 1 to 3 (DAK = 1)	67,420	130,084				67,420	87,458			
Months 1 to 3 (UIO = 1)	16,893	36,814				16,893	25,189			
Months 1 to 3 (UNV = 1)	54,769	110,145				54,769	74,191			
Months 1 to 3 (UWO = 1)	67,921	135,975				67,921	90,842			

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, to August 31, 2017, as of March 13, 2019.

Table C.4 (*continued*)

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

AAMC = Association of American Medical Colleges; ADA = University of California, San Diego; AMC = academic medical center; DAK = Dartmouth-Hitchcock; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; UIO = University of Iowa; UNV = University of Virginia; UWO = University of Wisconsin.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for the Association of American Medical Colleges (AAMC) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to AAMC. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on three core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for AAMC led to a Bayesian estimate of the program's impact on total Medicare expenditures of -2 percent (an estimated reduction of $20 per beneficiary per month) in the first quarter.

Table D.1. Comparison of frequentist and Bayesian impact estimates for AAMC in the first quarter after enrollment

	Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	-21 (-37, -5.9)	-20 (-53, 15)	-2%	-2%	-2%
Hospital admissions	-4.7 (-12, 2.7)	-5.0 (-15, 5.0)	-2%	-2%	-2%
ED visits	1.2 (-11, 13)	-12 (-31, 7.1)	-2%	< 1%	-2%

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, through August 31, 2017, as of March 13, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

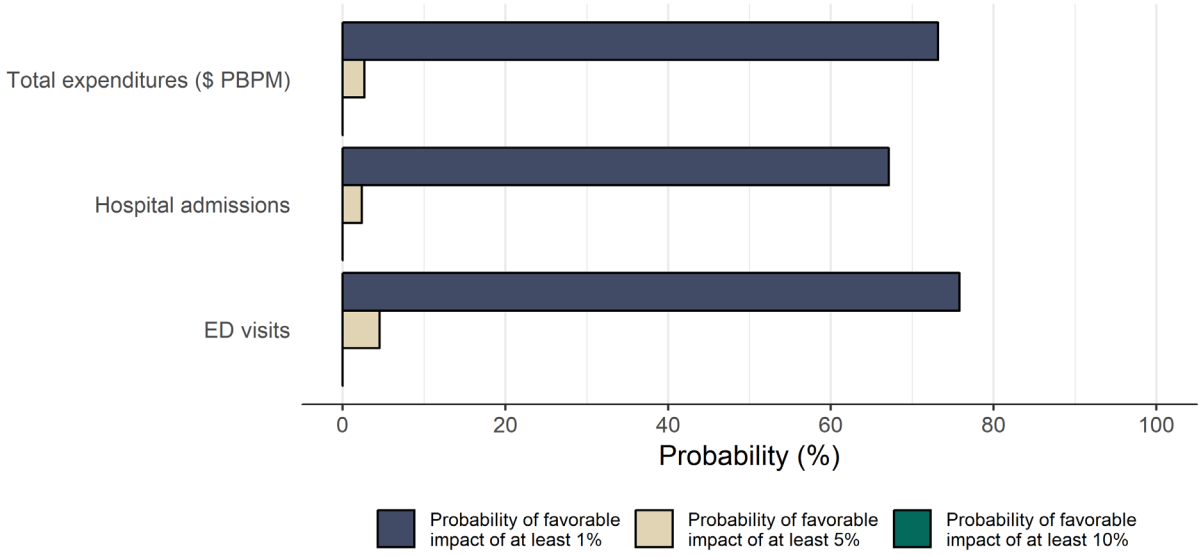
ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

The prior, based on the HCIA R1 data, coincided so completely with the frequentist impact estimates that the Bayesian and frequentist estimates for total expenditures and hospitalizations barely differ. However, the Bayesian estimate for ED visits differs noticeably from the frequentist estimate, for two reasons. First, the frequentist impact estimate for ED visits has the opposite sign from the estimates for expenditures and hospitalizations, and its uncertainty interval implies almost equal probabilities of favorable and unfavorable impacts. In the HCIA R1 data, impacts tend to be consistently favorable or unfavorable across outcomes for the same awardee, so give the uncertainty in the frequentist estimate, the Bayesian model brought the ED

visits impact estimate into line with the total expenditures and hospitalizations estimates. Second, in the HCIA R1 data impacts tend to be more pronounced immediately after the intervention begins and then diminish over time. For AAMC, the short follow-up period led the Bayesian model to estimate a larger impact on ED visits. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that AAMC achieved a favorable impacts in the first follow-up quarter on three core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the AAMC program had a favorable impact on key outcomes



Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2013, through August 31, 2017, as of March 13, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a moderate probability—in the range of 70 percent—that AAMC had a favorable impact of 1 percent or more on total Medicare expenditures, hospital admissions, and emergency department visits. However, the probability of a favorable impact of 5 percent or more is no more than 10 percent for any outcome. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the AAMC program did not have a meaningful impact on total expenditures or service utilization.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Avera Health

September 2020

Suhui (Evelyn) Li, Mynti Hossain, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

AVERA HEALTH

Avera Health (Avera), a nonprofit integrated health system, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the eLongTermCare (eLTC) program. The program offered a set of geriatric care and tele-health services to staff and residents in nursing facilities (NFs) across the Midwest. The goal of the program was to help NF residents gain access to timely, resident-centered care, particularly in rural areas that were geographically isolated from primary care physicians (PCPs) and geriatricians. The target population included all residents admitted to 45 NFs participating in the program. The program launched in November 2014 and the intervention period covered under HCIA R2 ended in August 2017. Table 1 summarizes the program's key characteristics.

Although Avera had experience providing tele-health services to providers and residents through an existing tele-health service model, its innovations in the eLTC program were (1) its method of providing instant, around-the-clock access to care specifically for NF residents through tele-health equipment; and (2) its creation and use of a risk-stratification algorithm for assessing the health risk of NF residents. Avera provided eLTC services out of a centrally staffed tele-health hub in Sioux Falls, South Dakota. Staff at the hub included clinicians, such as nursing staff and physicians, as well as support and administrative staff. Services included (1) staff training and empowerment, (2) tele-health transitional care coordination, and (3) tele-health consults for urgent and specialty care.

Important issues for understanding the evaluation

- The eLTC program aimed to help residents in NFs gain access to timely, resident-centered care, and thus reduce unnecessary transfers to emergency departments (EDs) and hospitals.
- The eLTC program represents a new component of Avera's existing tele-health service program that provided services such as ePharmacy and eEmergency.
- The impact analysis was based on 9,608 Medicare beneficiaries and 24,620 comparison beneficiaries with similar demographic and health characteristics who stayed in nonparticipating NFs in the same market locations, and thus were ineligible to participate in the program.
- The evaluation measured program impacts for long-term care and skilled care beneficiaries separately.

Due to the lack of regular presence of physicians at many nursing facilities, residents routinely experience long waiting periods to see PCPs or geriatricians, which can lead to costly and inconvenient transfers to EDs to receive timely medical attention. The awardee hypothesized that the eLTC program would lead NF staff to better identify situations that were truly emergent and to contact the eLTC team for nonemergent resident care issues. By training NF staff to use tele-health consults and providing such services to their residents, the program would better meet residents' medical needs and in turn reduce unnecessary transfers to EDs and hospitals, both of which would reduce the total cost of care.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Avera's eLTC program sought to help residents in NFs gain access to timely, resident-centered care, particularly in rural areas that were geographically isolated from PCPs and geriatricians, and thus reduce the number of unnecessary transfers to EDs and hospitals.
Major innovation	Avera's innovations were (1) providing instant, around-the-clock access to care for NF residents through tele-health equipment; and (2) creating and using a risk-stratification algorithm to assess the health risk of NF residents.
Program components	Quality improvement, telemedicine, and transitional care coordination
Target population	All residents at any of the NFs participating in the program
Participating providers	45 NFs across four states: Iowa, Minnesota, Nebraska, and South Dakota
Total enrollment	The eLTC program automatically enrolled all residents at the participating NFs. Avera reported that it passively enrolled 11,192 residents from November 2014 through August 2017, which represented 158 percent of its original enrollment target of 7,100 residents.
Level of engagement	Avera designed the eLTC program to be a NF-level intervention. It passively enrolled all residents at the participating NFs and did not focus on engaging them as program participants. Although residents could refuse tele-health consults if they did not want them, the awardee reported that residents infrequently declined services.
Theory of change or theory of action	The awardee hypothesized that the program lead NF staff to better identify situations that were truly emergent and to contact the eLTC team for nonemergent resident care issues. By training NF staff to use tele-health consults and providing such services to its residents, the program would better meet residents' medical needs and in turn reduce unnecessary transfers to EDs and hospitals, both of which would reduce the total cost of care.
Award amount	$8,827,572
Effective launch date	November 1, 2014
Program setting	NFs (provider based)
Market area	Rural, urban, and suburban
Market location	Iowa, Minnesota, Nebraska, and South Dakota
Target outcomes	• Reduced ED visits • Reduced hospitalizations • Lower total cost of care
Payment model	Capitated payment for services, based on a retail subscription model in which NFs make advance payments to use eLTC services for a specified month
Sustainability plans	Avera sustained the eLTC program at 53 NFs, including 33 of the sites that participated in the HCIA R2-funded program and an additional 20 sites that participated in the pilot of the payment model but did not participate in the award. Avera also continued to sell the program to other NFs and expanded its marketing efforts to include assisted living facilities.

ED = emergency department; eLTC = eLongTermCare; FFS = fee-for-service; HCIA R2 = Round 2 of the Health Care Innovation Awards; NF = nursing facility; and PCP = primary care physician.

The impact analysis presented in this report was based on treatment-eligible Medicare fee-for-service (FFS) beneficiaries, who stayed in participating NFs during the intervention period. The study sample included 9,608 treatment beneficiaries and 24,620 comparison beneficiaries with similar demographic and health characteristics who stayed in nonparticipating NFs in the same market locations as the participating facilities, and thus were ineligible to participate in the program. The evaluation measured program impacts separately for two types of beneficiaries: (1) long-term care residents with chronic conditions who needed assistance with daily activities and (2) skilled care patients who needed short-term skilled nursing services to recover from an acute medical condition. Table 2 summarizes the key features of the evaluation; Table A.1 in Appendix A describes the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group.
Intervention group for evaluation	The impact analysis relied on 9,608 Medicare FFS beneficiaries who stayed in participating NFs during the intervention period (86 percent of all enrollees). The analysis restricted the study sample to beneficiaries who could be linked to the Medicare enrollment data and who met the other study inclusion criteria, including enrolled in both Parts A and B and not enrolled in Medicare Advantage.
Comparison group	The impact analysis compares outcomes among participants to those of 24,620 comparison beneficiaries. The program selected the comparison group from beneficiaries with similar demographic and health characteristics who stayed in nonparticipating NFs in the same market locations as the participating facilities, and thus were ineligible to participate in the program.
Limitations	First, if there were unmeasured differences between participating NFs and nonparticipating NFs that correlated with the study outcomes, these differences could bias the results. Second, the evaluation was unable to capture the marginal benefit of tele-health transitional care coordination for high-risk beneficiaries due to the lack of data to identify the beneficiaries who received such services.

FFS = fee for service; NF = nursing facility.

PROGRAM DESIGN AND ADAPTATION

Avera's eLTC program had three components: (1) staff training and empowerment, (2) tele-health transitional care coordination, and (3) tele-health consults for urgent and specialty care.[1] Avera viewed the three components as equally important in reducing ED visits, hospitalizations, and total cost of care among NF residents.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Staff training and empowerment

Every month, Avera conducted one-hour trainings for all NF staff, either in person or virtually. Avera selected the topics based on program monitoring data and feedback from NF staff at each participating site. In the second program year, Avera began holding additional monthly meetings with NF staff to review any unplanned transfers to an ED or hospital that NFs made without first consulting the eLTC team. During these calls, Avera trained NF staff on how to identify (1) early changes in residents' conditions to avoid ED or hospital transfers and (2) situations that were truly emergent. Avera also provided informal training for NF staff through ad hoc calls and meetings throughout the program.

Tele-health transitional care coordination

Avera provided transitional care coordination services to newly admitted residents who were transferred from home, a hospital, or other care setting to a participating NF. First, the eLTC team assessed the health risk of new NF residents to differentiate between those who were at high versus low risk for ED and hospital transfers. Avera conducted manual reviews of all new residents to determine their risk level, while simultaneously developing its own risk-stratification algorithm. Second, depending upon a resident's risk level, the eLTC team at the tele-health hub delivered additional supports designed to improve care coordination. For high-risk residents, the eLTC team conducted a full geriatric evaluation and developed a tailored ePlan, which included a chronic disease management plan, a schedule for telephone and video consults, and a task list for NF staff to follow as appropriate. For low-risk residents, the eLTC team reviewed medication lists and provided medication recommendations to the PCP or an NF nurse. Upon the request of NF staff, low-risk residents could also receive a video consult or an ePlan.

Tele-health consults for urgent or specialty care

Avera provided around-the-clock tele-health consults for NF residents every day. Avera encouraged NF staff to call the eLTC team whenever a resident needed urgent medical care; the team member then evaluated the resident via direct two-way audio and video, if necessary, and instructed NF nurses on next steps for care. For nonurgent specialty care, the eLTC team worked with Avera specialists to schedule tele-health visits.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Over the three-year program, Avera successfully enrolled 45 NFs into the eLTC program. The eLTC program automatically enrolled all residents at the participating NFs. In turn, the awardee passively enrolled 11,192 residents, more than 150 percent of its original three-year projection.

Although Avera delivered the eLTC program at participating NFs largely as intended, the awardee experienced difficulty engaging providers with the use of tele-health consults, as some NF staff did not consistently initiate tele-health consults for their residents at the appropriate times. According to some NF administrators and staff, they did not use tele-health consults because they had work habits that were difficult to change or they were newly hired and thus not aware of the program's services. In some cases, residents' PCPs also reduced NF staff engagement with the program because PCPs were hesitant to allow NF staff to contact the eLTC team to serve their residents. In the survey of NF administrators and staff during the beginning of the third program year, more than two-thirds of respondents (68 percent) said that clinicians' resistance to the program was a barrier to meeting program goals.

Implications of program implementation for detecting impacts

- Because Avera implemented a facility-level intervention intended to improve the overall quality of care, impacts were measured over all residents, including those who did not directly receive program services.
- Avera's difficulty engaging providers in the eLTC tele-health consults might have limited the program's ability to achieve its desired outcomes.

ESTIMATING PROGRAM IMPACTS

Study sample

Because the eLTC program automatically enrolled all residents at the participating NFs, the treatment group for the impact study consisted of 9,608 Medicare FFS beneficiaries at participating NFs from November 2014 through August 2017. The awardee identified these eligible treatment beneficiaries using administrative data instead of its database, because the latter contained a significant proportion of records that could not be linked to Medicare data. Using administrative data also ensured the same process for identifying treatment and comparison groups, which is essential for obtaining a valid comparable group. A review of the awardee's database showed that more than 80 percent of the eligible treatment beneficiaries identified using administrative data enrolled in the eLTC program. The matched comparison group included 24,620 beneficiaries with similar demographic and health characteristics who stayed in a nonparticipating NFs in the same market locations as the participating facilities, and thus were ineligible to participate in the program. (Appendix A, Table A.1 describes the identification of the analytic sample.)

As noted previously, the eligible treatment sample included both long-term care residents and skilled care patients. Considering the differences in their health care needs, the analysis estimated the program's impact over the two groups of beneficiaries separately. The analysis defined skilled care beneficiaries as those who had a Medicare skilled nursing facility (SNF) claim that covered both the NF admission date and eLTC enrollment date. The analysis classified the remaining beneficiaries in the study sample as long-term care residents.

Characteristics of treatment and comparison group beneficiaries

A comparison of baseline characteristics confirmed that the treatment and comparison groups (including both long-term and skilled care beneficiaries) were well balanced (Table 3). The average age of treatment and comparison group beneficiaries during the baseline year was 80 years. Nearly two-thirds of both groups were male, and about 30 percent were dually eligible for Medicare and Medicaid. The average hierarchical condition category (HCC) risk score for both groups was 2.3, indicating that their predicted total Medicare expenditures were nearly two and a half times higher than the average for Medicare FFS beneficiaries nationally. Congestive heart failure and vascular disease were the most common chronic conditions among both groups. Appendix B provides the full balance results measured during the 12 months before enrollment.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

	Full group		Treatment group only	
Measure	Treatment (N = 9,608)	Comparison (N = 24,620)	Long-term care residents (N = 7,194)	Skilled care patients (N = 2,414)
Demographics				
Age at enrollment, years	80	80	81	80
Age group, %				
Younger than 65	8	9	8	8
65 to 74	18	20	17	21
75 to 84	32	31	32	34
85 and older	42	40	43	37
Male, %	65	65	66	62
White, %	94	93	95	93
Original reason for Medicare eligibility, %				
Old age and survivor's insurance	79	79	79	81
Disability insurance benefits[a]	20	20	20	18
Medicare/Medicaid dual status	30	30	33	18
Chronic conditions, %				
COPD	26	27	25	27
CHF	34	35	33	37
Morbid obesity	8	8	7	9
Vascular disease	31	29	33	25
Major depressive disorder	11	11	12	9.1
HCC score[b]				
Mean	2.3	2.4	2.2	2.7
25th percentile	1.2	1.2	1.1	1.5
Median	1.9	2.0	1.8	2.3
75th percentile	3.0	3.1	2.8	3.5
Case-mix index at admission[c]	42	42	40	49
Admitted after program launch, %	39	39	22	92

Table 3 (*continued*)

Measure	Full group		Treatment group only	
	Treatment (N = 9,608)	Comparison (N = 24,620)	Long-term care residents (N = 7,194)	Skilled care patients (N = 2,414)
Service use and expenditures during the year before enrollment				
Any hospitalizations, %	62	60	48	97
Any outpatient ED visits, %	46	45	43	51
Number of hospital admissions (per 1,000)	986	1,050	815	1,496
Number of outpatient ED visits (per 1,000)	753	741	706	892
Total Medicare expenditures ($ PBPM)	2,239	2,406	2,033	2,857

Source: Mathematica's analysis of Minimum Dataset and Medicare claims and enrollment data from November 2013 through February 2018, as of January 23, 2019.

Notes: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date depends upon whether the beneficiary was in a participating facility on the program start date. For an existing resident, the enrollment date is the date on which the program began at his or her facility. For a new resident admitted to the facility after the program launch, the enrollment date is the first day on which he or she became a resident in a participating facility.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

All beneficiary characteristics were measured during or as of the end of the baseline year. None of the differences between treatment and comparison groups in any of the baseline characteristics was statistically different from zero at the 0.10 level, two-tailed test.

Full balance results are presented in Appendix B. Exact matching variables include facility location, rural facility, and the number of days from a beneficiary's NF admission to enrollment in the program.

[a] Includes residents with both a disability and ESRD.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

[c] Case-mix index is a numeric score that reflects the relative resources predicted to provide care to a resident. A case-mix index is designated to each RUG under the CMS RUG-IV system. The higher the case-mix index, the greater the resource requirements for the resident.

CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disorder; CHF = congestive heart failure; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; PBPM = per beneficiary per month; RUG = resource utilization group.

Table 3 also shows that skilled care patients appeared to be sicker on average than long-term care residents. Skilled care patients had higher HCC scores at baseline and the program expected them to have higher needs for care during the NF stay (based on the case-mix index at NF admission) than long-term care residents. Skilled care patients also had significantly higher Medicare expenditures and hospital service use during the baseline year. Medicare coverage for nursing facility costs explains part of the difference in baseline expenditures and service use between the two groups of NF residents; Medicare covers skilled nursing care after a qualified acute hospital stay, but does not cover long-term nursing home care.

Analytic approach

The impact estimates are based on a difference-in-differences study design. This design measures program effects as the change in outcomes among eligible treatment beneficiaries before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Using an intent-to-treat sample minimized potential selection bias due to unobserved differences between the treatment and comparison groups. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. The primary outcomes are total Medicare spending, number of hospital admissions, and number of ED or observations visits. Secondary outcomes include the number of primary and specialty care visits.

The analysis defines the pre-enrollment period as the year before each resident's enrollment date and the post-enrollment period as the two years after. The enrollment date depends upon whether the beneficiary was in a participating facility on the program start date. For an existing resident at program launch, the enrollment date is the date on which the program began at his or her facility; for a new resident admitted to the facility after program launch, the enrollment date is the first day on which he or she became a resident in a participating facility. The analysis assigned a pseudo-enrollment date to each comparison beneficiary, based on an analogous method. Appendix A contains a detailed description of the statistical models used to estimate the effects of the program. Appendix C reports the full impact results. The program impacts for Avera were also estimated using a Bayesian approach, and presented in Appendix D.

IMPACT RESULTS

For long-term care residents, the program appears to have led to an estimated decrease in outpatient ED or observation visits and the probability of any hospitalization, both of which were statistically significant (Table 4). Although there was an estimated decrease in the number of hospital stays, especially in the second follow-up year, the estimate was not statistically significant. The estimated decrease in the number of primary care and specialty care visits were also small and not statistically significant, suggesting that the eLTC program had no discernible effect on physician visits among long-term care residents. The program's effects on service use led to a statistically significant reduction of 3.9 percent in total Medicare expenditures over the 24-month follow-up period. The estimated decrease in SNF expenditures accounted for about half of the reduction in total expenditures and was statistically significant for each year (Appendix C, Table C.3).

For skilled care beneficiaries, the estimated effect was large on ED and observation visits, representing a 10 percent reduction over the two-year period. There is also some evidence suggesting estimated small decreases in primary care and specialty care visits in the first follow-up year. However, these effects were not large enough to yield a statistically significant reduction in Medicare expenditures. Compared with long-term care beneficiaries, the lack of

program impact for skilled care beneficiaries could be due to less exposure to the intervention: given that skilled care beneficiaries on average spent fewer than 60 days in NFs, it is possible that they might have had less opportunity to receive the eLTC tele-health services.

Table 4. Estimated impact of the Avera eLTC program on selected outcomes

	Long-term care beneficiaries			Skilled care beneficiaries		
	Impact estimate	Percentage impact	*p*-value	Impact estimate	Percentage impact	*p*-value
Expenditures ($ PBPM)						
Year 1	-81*	-3.9%	0.07	83	3.2%	0.27
Year 2	-89*	-5.9%	0.09	-29	-2.0%	0.78
Cumulative	-73*	-3.9%	0.08	57	2.6%	0.44
Number of hospital stays, per 1,000 beneficiaries						
Year 1	-4.1	< 1%	0.84	-15	-2.9%	0.63
Year 2	-35	-6.1%	0.17	2.2	< 1%	0.96
Cumulative	-9.9	-1.4%	0.60	-8.8	-1.6%	0.77
Number of ED or observation visits, per 1,000 beneficiaries						
Year 1	-75***	-9.2%	< 0.01	-81*	-9.4%	0.06
Year 2	-66**	-9.3%	0.02	-108**	-14.2%	0.05
Cumulative	-73***	-9.3%	< 0.01	-85**	-10.1%	0.03
Number of primary care visits in ambulatory setting, per 1,000 beneficiaries						
Year 1	27	< 1%	0.74	-370**	-4.0%	0.02
Year 2	74	1.0%	0.47	88	1.3%	0.66
Cumulative	63	< 1%	0.42	-246	-2.9%	0.11
Number of specialty visits in any setting, per 1,000 beneficiaries						
Year 1	-76	< 1%	0.65	-609**	-5.8%	0.04
Year 2	-280	-3.2%	0.20	-449	-5.0%	0.28
Cumulative	-88	< 1%	0.59	-538*	-5.4%	0.07
Percentage of beneficiaries with any hospital admission						
Year 1	-1.3*	-3.0%	0.07	-0.8	-2.4%	0.52
Year 2	-3.1***	-8.4%	< 0.01	-0.8	-2.3%	0.67
Cumulative	-2.0***	-3.4%	< 0.01	-0.6	-1.2%	0.64
Sample sizes	**Treatment**	**Comparison**		**Treatment**	**Comparison**	
Year 1	7,194	19,713		2,414	4,907	
Year 2	4,486	13,724		1,200	2,479	

Source: Mathematica's analysis of Minimum Dataset and Medicare claims and enrollment data from November 2013 through February 2018, as of January 23, 2019.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between baseline and intervention periods. The impact estimates for the binary outcomes of any hospitalization and any readmission within 30 days of hospital discharge are regression-adjusted treatment–comparison differences based on a cross-sectional regression that controls for beneficiaries' characteristics and the outcome of interest at baseline. To account for extreme outliers in expenditures and number of visits or stays, the analysis trimmed outcome values for both groups at the 98th percentile of the treatment group distribution. A sensitivity analysis using the untrimmed outcome values shows that the

Table 4 (*continued*)

estimated impacts are similar. Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate. Appendix C presents full impact estimates. Appendix D shows the results from the Bayesian analysis.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; eLTC = eLongTermCare; PBPM = per beneficiary per month.

Three features of the eLTC program contributed to reduced ED and observation visits—and, for a more limited group, lower inpatient care use and Medicare expenditures—among eligible treatment beneficiaries relative to their matched comparison group. First, the awardee's experience implementing tele-health services under its existing tele-health service model facilitated implementing the eLTC program. According to Avera, its experience using tele-health technology and equipment enabled it to mitigate many technology and equipment challenges during eLTC program set-up.

Second, value-based payment policies for hospitals create an incentive for participating NFs to use the eLTC program. Hospitals with excess readmissions face penalties and payment reductions under Medicare's Hospital Readmissions and Reduction Program. Some participating NF administrators reported that hospitals in their referral network encouraged them to use tele-health consults when residents had urgent care needs as a way to help reduce hospital readmission.

Third, the quality of training and support that the awardee delivered to frontline NF staff encouraged using the tele-health consults. The awardee deployed a combination of formal training sessions and informal support meetings to ensure that all staff received full training on various resident care topics and on when and how to initiate tele-health consults; the awardee ultimately trained 2,167 NF staff—about 105 percent of its three-year training goal. In addition, the awardee carried out ongoing outreach efforts to PCPs and NF staff to encourage them to increase their use of tele-health services. As a result of these efforts, most staff who responded to the survey at the start of the third program year believed that the eLTC program had positive impacts on the delivery of care and health outcomes.

Main findings from impact evaluation

- The eLTC program led to reduced total Medicare expenditures, ED or observation visits, and the probability of any hospitalization for long-term care beneficiaries in participating NFs.
- The program also led to a reduction in ED or observation visits for skilled-care beneficiaries. But there was no discernable program effect on Medicare expenditures for those beneficiaries.

CONCLUSION

Overall, Avera was mostly successful in implementing the eLTC program to provide tele-health services to staff and residents at participating NFs. The program achieved the goal of reducing ED visits among NF residents. Particularly for long-term care residents, the program also led to reduced Medicare expenditures and the likelihood of hospital admission. The latter finding suggests that even though Medicare does not cover most long-term care, implementing the program might generate savings for Medicare through reduced hospital and ED use among Medicare beneficiaries residing in nursing facilities.

Limitations of evaluation

The analysis has several limitations. First, it selected all matched comparison beneficiaries from nonparticipating NFs. Although the matching algorithm took facility characteristics into account, and the residents in the two groups were well-matched, if there were unmeasured differences between participating NFs and nonparticipating NFs that correlated with the study outcomes, these differences could bias the results. Second, the eLTC program provided tailored tele-health transitional care coordination services to high-risk residents, which might have been important for achieving favorable outcomes and for replicating in other settings. However, the evaluation was unable to capture the marginal benefit of tele-health transitional care coordination for high-risk beneficiaries due to the lack of data to identify the beneficiaries who received such services.

PROGRAM SUSTAINABILITY

After its award ended in August 2017, Avera sustained the eLTC program at 53 NFs, continued to sell the program to other NFs, and expanded its marketing efforts to include assisted living facilities. The 53 sites included 33 of the sites that participated during the award period, and an additional 20 sites that participated in the pilot of the payment model but did not participate in the award.Avera implemented a retail subscription payment model to cover the cost of the program. Under this model, NFs pay Avera a monthly lump-sum payment to have access to its eLTC services. The NFs expect that eLTC services will help them reduce avoidable ED visits and hospital admissions, which

Avera's proposed payment models

Avera has a retail subscription payment model in which NFs make two types of payments to Avera for every resident in the NF's monthly census, regardless of payer or length of stay: (1) an initial fee at the resident's time of admission and (2) a monthly fee until the resident is discharged. Services covered include the following:

- NF staff training and empowerment
- Tele-health transitional care coordination
- Urgent and specialty care consults via Avera's tele-health hub

Avera also proposed a Medicare payment model that was performance-based with two-sided risk.

they can then leverage to partner with accountable care organizations or other entities pursuing value-based purchasing arrangements with the goal of lowering Medicare costs.

In addition, Avera had proposed a Medicare payment model that was performance-based with two-sided risk. Payments could be reduced by up to 50 percent for failing to meet benchmarks on certain utilization and quality measures, such as short-stay rehospitalization rate, ED transfer rate, or 24/7 access to geriatric care. The awardee reported that it had proposed the payment model to the Physician-Focused Payment Model Technical Advisory Committee, which had recommended it to the Secretary of Health and Human Services. However, the Secretary decided not to pursue the model.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits/stays are based on a difference-in-differences approach with beneficiary fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or emergency department (ED) visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or any ED visit during the baseline period. The intervention years are beneficiary-specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of outcomes.

The evaluation measured program impacts separately for two types of beneficiaries: (1) long-term care residents with chronic conditions who needed assistance with daily activities and (2) skilled care patients who needed short-term skilled nursing services to recover from an acute medical condition.

Table A.1 shows the how the analytic sample for this study was defined. It lists the reasons why participants were excluded and the number of participants withdrawn for each reason.

Table A.1. Identification of final sample for impact analysis for Avera

	Number of beneficiaries removed from analytic sample	Number of beneficiaries remaining in analytic sample
Total beneficiary records found in MDS through August 31, 2017		**26,459**
Repeated records of unique beneficiaries	5,921	20,538
Beneficiaries who:		
Were not found in Medicare enrollment database	541	19,997
Were not alive at the time of program launch	6,524	13,473
Were not enrolled in both Medicare Part A and Part B	365	13,108
Were enrolled in Medicare Advantage	2,870	10,238
Did not have Medicare as primary payer	57	10,181
Lacked 90 days of FFS enrollment during baseline	68	10,113
Died within 30 days of enrollment	505	9,608
Final analytic sample		**9,608**

Source: Mathematica's analysis of Minimum Dataset and Medicare claims and enrollment data from November 2013 through February 2018, as of January 23, 2019.

FFS = fee-for-service; MDS = Minimum Dataset.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 show the variables used for matching for long-term care residents. The table displays the weighted means of baseline characteristics for the 7,194 treatment beneficiaries and the 19,713 matched comparison beneficiaries used in the impact analysis. Table B.2 shows the same information for 2,414 skilled-care treatment beneficiaries and their 4,907 matched comparison beneficiaries. Both tables show the means, difference in means, the percentage difference, and the standardized difference for each variable, calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables include demographic characteristics (age, gender, and race); Medicare dual eligibility status; health status (as measured by the hierarchical condition category [HCC] score and chronic condition indicators); Medicare expenditures; service use, and facility-level characteristics. The analysis required an exact match on facility location, rural facility status, and the number of days between a beneficiary's nursing facility (NF) admission and program enrollment or pseudo enrollment (0, 1 to 180, and 180 or more days). The variables are measured over various specified intervals within the 12 months before enrollment in the intervention.

The tables also show the results of the equivalency-of-means tests. p-values come from a weighted two-sample t-test, which tests for whether the difference in the means is statistically significant. The equivalence test p-values are the greater of two one-sided weighted t-test p-values equivalence test, which assesses whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the study team performed an omnibus test in which the null hypothesis is that the treatment and matched comparison groups are balanced across all linear combinations of the covariates. The team used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

Tables B.1 and B.2 show that the omnibus test rejects the joint hypothesis that the treatment and comparison groups are equal on all baseline characteristics. However, for the great majority of variables, the differences in treatment and comparison means are small and not statistically significant. For a few variables with statistically significant differences in means, such as the baseline total expenditures, the absolute differences are small and the standard differences are less than 10 percent in most cases. Taken together, the results suggest that the treatment and comparison groups are well balanced. The regression model accounts for the remaining differences between the two groups through the difference-in-differences approach and control variables.

Table B.1. Baseline characteristics of treatment and matched comparison groups for long-term care beneficiaries

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	81 (0.13)	81 (0.08)	0.35 (0.19)	< +/-1	0.03	0.06	< 0.01
Female, %	66 (0.56)	65 (0.34)	0.78 (0.79)	1.2	0.02	0.32	< 0.01
White, %	95 (0.27)	94 (0.14)	0.85 (0.40)	< +/-1	0.04	0.03	< 0.01
Number of days between NF admission and enrollment, %							
Days between admission and enrollment: 0	22 (0.49)	22 (0.31)	0.06 (0.66)	< +/-1	0.00	0.93	< 0.01
Days between admission and enrollment: 1 to 180	11 (0.36)	11 (0.22)	-0.01 (0.52)	< +/-1	0.00	0.98	< 0.01
Days between admission and enrollment: 181 or more	68 (0.55)	68 (0.34)	-0.04 (0.75)	< +/-1	0.00	0.96	< 0.01
Medicare dual eligibility status, %							
Dually eligible for Medicare and Medicaid	33 (0.56)	34 (0.33)	-0.23 (0.82)	< +/-1	0.00	0.78	< 0.01
Health status and diagnoses							
HCC score[a]	2.22 (0.02)	2.30 (0.01)	-0.08 (0.03)	-3.8	-0.05	< 0.01	< 0.01
Vascular disease, %	33 (0.55)	30 (0.32)	2.1 (0.78)	6.4	0.04	< 0.01	< 0.01
COPD, %	25 (0.51)	27 (0.31)	-1.3 (0.70)	-4.9	-0.03	0.07	< 0.01
Diabetes with acute complications, %	0.70 (0.10)	0.76 (0.06)	-0.07 (0.15)	-9.7	-0.01	0.64	< 0.01
Morbid obesity, %	7.2 (0.31)	7.9 (0.18)	-0.70 (0.45)	-9.7	-0.03	0.12	< 0.01
Major depressive disorder, %	12 (0.38)	11 (0.22)	0.33 (0.55)	2.8	0.01	0.55	< 0.01
CHF, %	33 (0.55)	34 (0.33)	-1.8 (0.79)	-5.6	-0.04	0.02	< 0.01
Medicare expenditures							
Total expenditures	2,060 (35)	2,207 (18)	-148 (49)	-7.2	-0.05	< 0.01	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total expenditures, 7 days before enrollment	5,858 (226)	4,661 (112)	1,197 (270)	20	0.06	< 0.01	< 0.01
Total expenditures, 3 months before enrollment	2,737 (69)	2,802 (31)	-65 (91)	-2.4	-0.01	0.48	< 0.01
Hospice expenditures	49 (4.2)	40 (2.0)	8.8 (5.7)	18	0.02	0.12	< 0.01
Service use							
Total hospitalizations	824 (14)	898 (8.1)	-75 (20)	-9.1	-0.06	< 0.01	< 0.01
Total hospitalizations, 3 months before enrollment	1,221 (29)	1,337 (17)	-116 (41)	-9.5	-0.05	< 0.01	< 0.01
Total ED or observation visits	871 (20)	849 (9.8)	21 (26)	2.5	0.01	0.40	< 0.01
Total ED or observation visits, 3 months before enrollment	1,100 (32)	1,089 (18)	10 (43)	< +/-1	0.00	0.81	< 0.01
Primary care visits in any setting	10,771 (107)	11,186 (60)	-415 (151)	-3.9	-0.05	< 0.01	< 0.01
Primary care visits in any setting, 3 months before enrollment	13,010 (191)	13,640 (105)	-630 (266)	-4.8	-0.04	0.02	< 0.01
Hospice use, %	2.1 (0.17)	1.8 (0.09)	0.29 (0.23)	14	0.02	0.21	< 0.01
RUG-IV group: rehabilitation, %	73 (0.52)	74 (0.32)	-0.78 (0.73)	-1.1	-0.02	0.28	< 0.01
RUG-IV group: reduced physical function, %	13 (0.40)	13 (0.25)	0.74 (0.56)	5.6	0.02	0.18	< 0.01
RUG-IV group: clinically complex, %	4.8 (0.25)	5.1 (0.16)	-0.29 (0.36)	-6.0	-0.01	0.42	< 0.01
CMI score	40 (0.22)	40 (0.13)	-0.39 (0.30)	< +/-1	-0.02	0.20	< 0.01
Facility-level factors							
Nonprofit facility, %	38 (0.57)	46 (0.35)	-7.6 (0.84)	-20	-0.16	< 0.01	< 0.01
Number of federally certified beds	93 (0.40)	96 (0.34)	-2.6 (0.74)	-2.8	-0.08	< 0.01	< 0.01
Overall 5-star rating	3.1 (0.01)	3.1 (0.01)	-0.04 (0.02)	-1.1	-0.03	0.07	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Facility hospice expenditures for residents with hospice use	7,083 (33)	7,125 (20)	-41 (46)	< +/-1	-0.01	0.37	< 0.01
Percentage of facility residents with Medicare hospice expenditures	0.11 (0.00)	0.10 (0.00)	0.00 (0.00)	3.6	0.09	< 0.01	< 0.01
Propensity score	0.06 (0.00)	0.06 (0.00)	0.00 (0.00)	4.4	0.09	< 0.01	< 0.01
Number of beneficiaries	7,194	19,713					
Omnibus test				Chi-squared statistic 581.47	Degrees of freedom 34.00	P-value 0.00	

Source: Mathematica's analysis of Minimum Dataset and Medicare claims and enrollment data from November 2013 through February 2018, as of January 23, 2019. Facility characteristics derived from CMS Provider of Services and Nursing Home Compare data.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include facility location, rural facility, and the number of days from a beneficiary's NF admission to enrollment in the program.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CHF = congestive heart failure; CMI = case mix index; COPD = chronic obstructive pulmonary disorder; ED = emergency department; HCC = hierarchical condition category; NF = nursing facility; PBPM = per beneficiary per month; RUG = resource utilization group; SE = standard error.

Table B.2. Baseline characteristics of treatment and matched comparison groups for skilled care beneficiaries

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	79 (0.23)	78 (0.16)	1.1 (0.33)	1.4	0.10	< 0.01	< 0.01
Female, %	62 (0.99)	62 (0.69)	-0.35 (1.4)	< +/-1	-0.01	0.80	< 0.01
White, %	93 (0.51)	91 (0.34)	2.3 (0.76)	2.5	0.09	< 0.01	< 0.01
Medicare dual eligibility status, %							
Dually eligible for Medicare and Medicaid	18 (0.79)	20 (0.56)	-2.3 (1.1)	-12	-0.06	0.04	< 0.01
Health status and diagnosis							
HCC score[a]	2.66 (0.03)	2.70 (0.02)	-0.04 (0.05)	-1.7	-0.03	0.37	< 0.01
Vascular disease, %	25 (0.88)	25 (0.61)	0.06 (1.3)	< +/-1	0.00	0.97	< 0.01
COPD, %	27 (0.90)	27 (0.63)	-0.35 (1.3)	-1.3	-0.01	0.79	< 0.01
Diabetes with acute complications, %	1.2 (0.22)	1.6 (0.17)	-0.46 (0.33)	-39	-0.04	0.17	< 0.01
Morbid obesity, %	9.2 (0.59)	9.7 (0.43)	-0.55 (0.82)	-6.0	-0.02	0.50	< 0.01
Major depressive disorder, %	9.1 (0.59)	9.7 (0.43)	-0.61 (0.82)	-6.7	-0.02	0.46	< 0.01
CHF, %	37 (0.98)	36 (0.69)	0.73 (1.4)	2.0	0.02	0.60	< 0.01
Medicare expenditures							
Total expenditures	2,912 (58)	3,065 (39)	-153 (82)	-5.2	-0.05	0.06	< 0.01
Total expenditures, 7 days before enrollment	35,813 (751)	32,698 (488)	3,115 (1,089)	8.7	0.08	< 0.01	< 0.01
Total expenditures, 3 months before enrollment	7,822 (150)	8,176 (103)	-354 (212)	-4.5	-0.05	0.09	< 0.01
Hospice expenditures	4.0 (1.7)	2.0 (0.58)	2.0 (2.1)	49	0.02	0.34	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Service use							
Total hospitalizations	1,533 (28)	1,536 (15)	-2.5 (36)	< +/-1	0.00	0.94	< 0.01
Total hospitalizations, 3 months before enrollment	4,549 (50)	4,566 (31)	-17 (69)	< +/-1	-0.01	0.80	< 0.01
Total ED or observation visits	1,064 (38)	1,019 (23)	45 (49)	4.2	0.02	0.36	< 0.01
Total ED or observation visits, 3 months before enrollment	1,892 (72)	1,737 (46)	155 (96)	8.2	0.04	0.10	< 0.01
Primary care visits in any setting	12,266 (201)	12,498 (134)	-232 (280)	-1.9	-0.02	0.41	< 0.01
Primary care visits in any setting, 3 months before enrollment	24,491 (424)	25,162 (315)	-672 (645)	-2.7	-0.03	0.30	< 0.01
Hospice use, %	0.17 (0.08)	0.06 (0.03)	0.11 (0.10)	67	0.03	0.25	< 0.01
RUG-IV group: rehabilitation, %	95 (0.46)	94 (0.35)	0.60 (0.67)	< +/-1	0.03	0.37	< 0.01
RUG-IV group: reduced physical function, %	1.8 (0.27)	1.9 (0.19)	-0.09 (0.40)	-5.0	-0.01	0.82	< 0.01
RUG-IV group: clinically complex, %	2.2 (0.30)	2.4 (0.23)	-0.24 (0.43)	-11	-0.02	0.57	< 0.01
CMI score	49 (0.22)	49 (0.16)	0.67 (0.31)	1.4	0.06	0.03	< 0.01
Facility-level factors							
Nonprofit facility, %	34 (0.96)	36 (0.68)	-2.7 (1.4)	-8.0	-0.06	0.05	< 0.01
Number of federally certified beds	94 (0.67)	90 (0.65)	4.0 (1.2)	4.2	0.12	< 0.01	< 0.01
Overall 5-star rating	3.1 (0.02)	3.1 (0.02)	-0.06 (0.03)	-2.1	-0.05	0.06	< 0.01
Facility hospice expenditures for residents with hospice use	6,973 (57)	7,344 (38)	-371 (84)	-5.3	-0.13	< 0.01	< 0.01
Percentage of facility residents with Medicare hospice expenditures	0.10 (0.00)	0.11 (0.00)	-0.01 (0.00)	-6.0	-0.16	< 0.01	< 0.01
Propensity score	0.11 (0.00)	0.09 (0.00)	0.01 (0.00)	13	0.18	< 0.01	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Number of beneficiaries	2,414	4,907					
Omnibus test				Chi-squared statistic 597.50	Degrees of freedom 32.00	P-value 0.00	

Source: Mathematica's analysis of Minimum Dataset and Medicare claims and enrollment data from November 2013 through February 2018, as of January 23, 2019. Facility characteristics derived from CMS Provider of Services and Nursing Home Compare data.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. p-values come from a weighted two-sample t-test; equivalence test p-values are the greater of the p-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include facility location, rural facility, and the number of days from a beneficiary's NF admission to enrollment in the program. ED visit measures include observation stays.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CHF = congestive heart failure; CMI = case mix index; COPD = chronic obstructive pulmonary disorder; ED = emergency department; HCC = hierarchical condition category; NF = nursing facility; PBPM = per beneficiary per month; RUG = resource utilization group; SE = standard error.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Tables C.1 and C.2 display the results from the impact analysis. Tables C.1 and C.2 show the impact estimates for long-term care beneficiaries and skilled care beneficiaries, respectively. The analysis measured all outcomes separately over intervention Years 1 and 2. It estimated the models over Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. Impact estimates that are statistically different from zero at the .10, .05, and .01 levels, using a two-tailed test, are indicated with one, two, and three asterisks, respectively.

Table C.1. Estimated impact of the Avera eLTC intervention on selected Medicare FFS expenditures (dollars PBPM) and use measures during 1- and 2-year follow-up periods: Long-term care beneficiaries

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)					
Baseline year	1,987	2,148			
Year 1	2,109	2,351	-81* (44)	-3.9%	0.07
Year 2	1,855	2,105	-89* (53)	-5.9%	0.09
Cumulative	1,970	2,204	-73* (42)	-3.9%	0.08
Acute inpatient expenditures ($ PBPM)					
Baseline year	737	807			
Year 1	705	774	1.9 (22)	< 1%	0.93
Year 2	633	728	-24 (27)	-5.1%	0.37
Cumulative	661	734	-2.6 (21)	< 1%	0.90
Hospital outpatient expenditures ($ PBPM)					
Baseline year	334	326			
Year 1	362	360	-6.2 (6.9)	-1.7%	0.37
Year 2	345	334	3.0 (9.2)	< 1%	0.75
Cumulative	356	351	-2.7 (6.8)	< 1%	0.69
Professional Part B expenditures ($ PBPM)					
Baseline year	266	289			
Year 1	251	281	-7.4* (4.3)	-3.0%	0.08
Year 2	233	268	-12** (5.3)	-5.8%	0.03
Cumulative	241	272	-8.5** (4.1)	-3.6%	0.04
SNF expenditures ($ PBPM)					
Baseline year	387	423			
Year 1	473	557	-48** (19)	-9.7%	0.01
Year 2	290	364	-39** (20)	-14.5%	0.05
Cumulative	405	480	-39** (17)	-9.2%	0.02
Hospital stays, per 1,000 beneficiaries					
Baseline year	805	888			

Table C.1 (*continued*)

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Year 1	793	880	-4.1 (20)	< 1%	0.84
Year 2	738	855	-35 (25)	-6.1%	0.17
Cumulative	753	845	-9.9 (19)	-1.4%	0.60
ED or observation visits, per 1,000 beneficiaries					
Baseline year	814	831			
Year 1	779	871	-75*** (22)	-9.2%	< 0.01
Year 2	757	841	-66** (28)	-9.3%	0.02
Cumulative	772	863	-73**** (21)	-9.3%	< 0.01
Percentage of beneficiaries with any hospital admission in a time period					
Baseline year	48	50			
Year 1	42	43	-1.3* (0.70)	-3.0%	0.07
Year 2	33	37	-3.1*** (0.98)	-8.4%	< 0.01
Cumulative	58	59	-2.0*** (0.73)	-3.4%	< 0.01
Percentage of beneficiaries with any ED or observation visits in a time period					
Baseline year	43	45			
Year 1	41	44	-2.7*** (0.74)	-6.1%	< 0.01
Year 2	38	40	-1.6 (1.00)	-4.1%	0.10
Cumulative	58	61	-2.9**** (0.75)	-4.7%	< 0.01
Primary care visits in ambulatory setting, per 1,000 beneficiaries					
Baseline year	7,481	7,931			
Year 1	8,154	8,576	27 (82)	< 1%	0.74
Year 2	7,683	8,059	74 (101)	1.0%	0.47
Cumulative	7,971	8,358	63 (78)	< 1%	0.42
Specialist visits in all settings, per 1,000 beneficiaries					
Baseline year	10,288	10,560			
Year 1	10,043	10,390	-76 (165)	< 1%	0.65
Year 2	9,362	9,915	-280 (219)	-3.2%	0.20
Cumulative	9,739	10,099	-88 (165)	< 1%	0.59
Percentage of hospital discharges with a 30-day readmission					
Baseline year	18	18			
Year 1	22	22	-1.0 (1.0)	-4.5%	0.30
Year 2	18	18	-0.89 (1.4)	-4.7%	0.52
Cumulative	21	21	-0.98 (0.94)	-4.5%	0.30
Sample sizes					
Number of beneficiaries					
Baseline year	7,194	19,713			
Year 1	7,194	19,713			
Year 2	4,486	13,724			
Cumulative	7,194	19,713			

Table C.1 (*continued*)

Source: Mathematica's analysis of information from Minimum Data Set and Medicare claims and enrollment data from November 2013 through February 2018, as of January 23, 2019.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary-specific and defined relative to each beneficiary's date of enrollment. To account for extreme outliers in expenditures and number of visits or stays, the analysis trimmed the outcome values for both groups at the 98th percentile of the treatment group distribution. The analysis determined 98th percentile values for top-coding from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and the follow-up years.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; eLTC = eLongTermCare; FFS = fee-for-service; PBPM = per beneficiary per month; SNF = skilled nursing facility.

Table C.2. Estimated impact of the Avera eLTC program on selected Medicare FFS expenditures (dollars PBPM) and use measures during 1- and 2-year follow-up periods: Skilled care beneficiaries

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)					
Baseline year	2,810	2,975			
Year 1	2,823	2,906	83 (75)	3.2%	0.27
Year 2	1,932	2,126	-29 (102)	-2.0%	0.78
Cumulative	2,493	2,602	57 (74)	2.6%	0.44
Acute inpatient expenditures ($ PBPM)					
Baseline year	1,501	1,574			
Year 1	505	525	52 (40)	12.5%	0.19
Year 2	644	668	48 (56)	10.8%	0.39
Cumulative	524	537	60 (40)	13.9%	0.13
Hospital outpatient expenditures ($ PBPM)					
Baseline year	369	346			
Year 1	418	396	-1.2 (15)	< 1%	0.93
Year 2	402	363	16 (20)	4.9%	0.41
Cumulative	418	385	8.9 (15)	2.3%	0.54
Professional Part B expenditures ($ PBPM)					
Baseline year	378	416			
Year 1	252	291	-2.0 (7.9)	< 1%	0.80
Year 2	241	279	-1.4 (11)	< 1%	0.90
Cumulative	244	284	-2.5 (7.8)	-1.1%	0.75

Table C.2 (*continued*)

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
SNF expenditures ($ PBPM)					
Baseline year	292	303			
Year 1	1,225	1,215	22 (37)	1.9%	0.55
Year 2	337	400	-52 (37)	-20.0%	0.16
Cumulative	947	960	-1.9 (34)	< 1%	0.96
Hospital stays, per 1,000 beneficiaries					
Baseline year	1,472	1,507			
Year 1	567	617	-15 (32)	-2.9%	0.63
Year 2	705	738	2.2 (47)	< 1%	0.96
Cumulative	592	636	-8.8 (30)	-1.6%	0.77
ED or observation visits, per 1,000 beneficiaries					
Baseline year	97	97			
Year 1	815	896	-81* (43)	-9.4%	0.06
Year 2	729	837	-108** (55)	-14.2%	0.05
Cumulative	793	878	-85** (40)	-10.1%	0.03
Percentage of beneficiaries with any hospital admission in a time period					
Baseline year	97	97			
Year 1	33	34	-0.80 (1.3)	-2.4%	0.52
Year 2	34	35	-0.79 (1.9)	-2.3%	0.67
Cumulative	52	52	-0.65 (1.4)	-1.2%	0.64
Percentage of beneficiaries with any ED or observation visits in a time period					
Baseline year	51	50			
Year 1	43	44	-1.4 (1.3)	-3.1%	0.30
Year 2	37	39	-2.1 (1.9)	-5.4%	0.27
Cumulative	59	62	-2.8** (1.3)	-4.5%	0.04
Primary care visits in ambulatory setting, per 1,000 beneficiaries					
Baseline year	6,827	7,096			
Year 1	9,131	9,771	-370** (157)	-4.0%	0.02
Year 2	7,385	7,567	88 (201)	1.3%	0.66
Cumulative	8,564	9,080	-246 (153)	-2.9%	0.11
Primary care visits in all settings, per 1,000 beneficiaries					
Baseline year	11,954	12,242			
Year 1	11,760	12,476	-428* (243)	-3.7%	0.08
Year 2	10,178	10,655	-189 (331)	-2.1%	0.57
Cumulative	11,111	11,804	-405* (238)	-3.7%	0.09
Specialist visits in all settings, per 1,000 beneficiaries					
Baseline year	13,156	13,209			
Year 1	10,228	10,890	-609** (300)	-5.8%	0.04
Year 2	9,657	10,158	-449 (418)	-5.0%	0.28
Cumulative	9,923	10,514	-538* (296)	-5.4%	0.07

Table C.2 (*continued*)

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Percentage of hospital discharges with a 30-day readmission					
Baseline year	21	19			
Year 1	13	13	-0.52 (1.5)	-4.0%	0.73
Year 2	17	12	4.7** (2.3)	38.6%	0.05
Cumulative	13	13	0.26 (1.5)	2.0%	0.86
Sample sizes					
Number of beneficiaries					
Baseline year	2,414	4,907			
Year 1	2,414	4,907			
Year 2	1,200	2,479			
Cumulative	2,414	4,907			

Sources: Mathematica's analysis of information from Minimum Data Set and Medicare claims and enrollment data from November 2013 through February 2018, as of January 23, 2019.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary-specific and defined relative to each beneficiary's date of enrollment. To account for extreme outliers in expenditures and number of visits or stays, the analysis trimmed the outcome values for both groups at the 98th percentile of the treatment group distribution. The analysis determined 98th percentile values for top-coding from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

**** Significantly different from zero at the .001 level, two-tailed test.

ED = emergency department; eLTC = eLongTermCare; FFS = fee-for-service; PBPM = per beneficiary per month; SNF = skilled nursing facility.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for Avera were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to Avera. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts for long-term care and skilled care beneficiaries on CMS's four core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for CMS's four core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regressions for Avera led to a Bayesian estimate of the program's impact on total Medicare expenditures of between -3 and -4 percent (an estimated reduction of between $63 and $88 per beneficiary per month) for long-term care beneficiaries in the first two years and an impact of less than 1 percent (an estimated increase of between $5 and $19 per beneficiary per month) for skilled care beneficiaries in the same period.

Table D.1. Comparison of frequentist and Bayesian impact estimates for Avera in the first two years after enrollment

Subgroup	Outcome	Follow-up period	Impact estimate (95 percent interval)		Percentage impacts		
			Frequentist	Bayesian	Prior	Frequentist	Bayesian
Long-term care	Total expenditures ($ PBPM)	Year 1	-81 (-168, 5.9)	-88 (-155, -23)	> -1%	-4%	-4%
		Year 2	-89 (-192, 15)	-63 (-120, -5.9)	< 1%	-5%	-3%
	Hospital admissions	Year 1	-4.1 (-44, 36)	-30 (-56, -6.4)	< 1%	> -1%	-4%
		Year 2	-35 (-85, 15)	-24 (-48, -0.23)	< 1%	-5%	-3%
	ED visits	Year 1	-75 (-118, -32)	-37 (-67, -11)	> -1%	-9%	-4%
		Year 2	-66 (-120, -12)	-30 (-58, -3.9)	< 1%	-8%	-4%
	Readmissions	Year 1	-0.01 (-0.03, 0.01)	-0.01 (-0.02, 0.00)	< 1%	-5%	-4%
		Year 2	-0.01 (-0.04, 0.02)	-0.01 (-0.01, 0.00)	< 1%	-5%	-3%
Skilled care	Total expenditures ($ PBPM)	Year 1	83 (-65, 230)	5.0 (-110, 121)	< 1%	3%	< 1%
		Year 2	-29 (-229, 171)	19 (-63, 100)	< 1%	-1%	< 1%
	Hospital admissions	Year 1	-15 (-77, 47)	2.2 (-23, 27)	< 1%	-3%	< 1%
		Year 2	2.2 (-90, 94)	8.1 (-22, 39)	1%	< 1%	1%
	ED visits	Year 1	-81 (-165, 2.2)	-1.5 (-44, 38)	> -1%	-9%	> -1%
		Year 2	-108 (-215, -1.0)	5.1 (-34, 41)	< 1%	-13%	< 1%
	Readmissions	Year 1	-0.01 (-0.04, 0.02)	0.00 (-0.01, 0.01)	< 1%	-4%	< 1%
		Year 2	0.05 (0.00, 0.09)	0.00 (0.00, 0.01)	< 1%	39%	1%

Source: Mathematica's analysis of Medicare claims as of August 31, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. Readmissions impacts are calculated as the percentage of discharges with a 30-day readmission and are presented on a 0-1 scale where 0.05 represents 5 percent. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

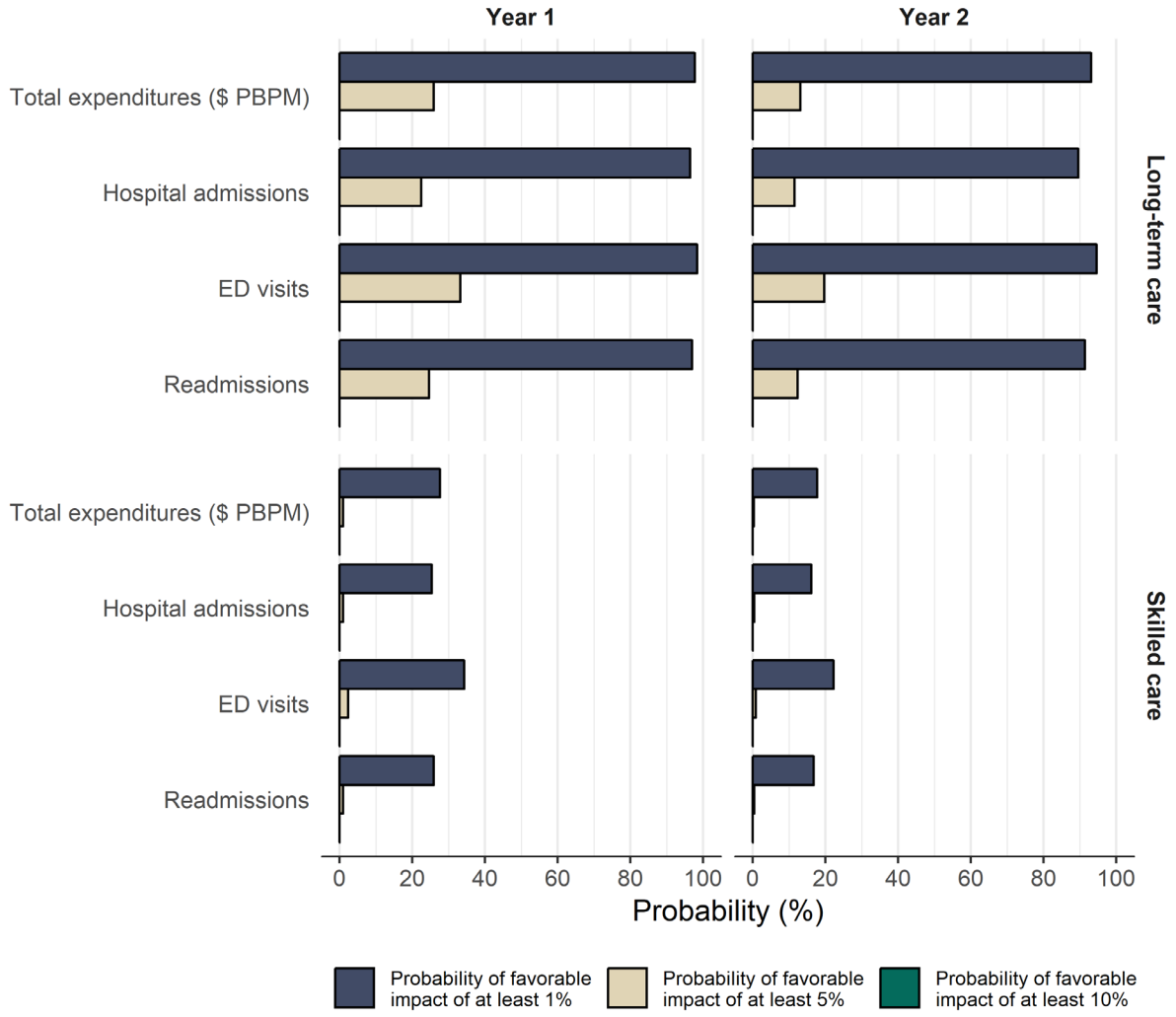
ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results are imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that for long-term care beneficiaries, Avera reduced Medicare expenditures and ED visit rates in both of the first two years. Additionally, the Bayesian analysis suggests that Avera reduced hospital admissions in each of the first two years. For skilled care beneficiaries, the Bayesian results also corroborate the frequentist finding that impacts are generally statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that Avera achieved favorable impacts for each sample during each of the first two years on CMS's four core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

As the figure shows, there is a strong probability—in the range of 90-100 percent—that Avera had a favorable impact of 1 percent or more on all four core outcomes in the first two program years for long-term care beneficiaries. However, the probability of an impact of 5 percent or more is more modest, at approximately 20 percent in the first year and 10 percent in the second year, suggesting that there is a high probability of a small impact for the long-term care beneficiary sample. Impact probabilities are much lower in the skilled care beneficiary sample, where the likelihood of an impact of 1 percent or more is close to 30 percent for all four outcomes in the first year and 20 percent in the second. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the Avera program appears promising for long-term care beneficiaries, although its impacts are small and do not extend to skilled care beneficiaries.

Figure D.1. Probability that the Avera program had a favorable impact on key outcomes



Source: Mathematica's analysis of Medicare claims as of August 31, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. Readmissions impacts are calculated as the percentage of discharges with a 30-day readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: The Board of Trustees at The University of Illinois, Chicago

September 2020

Kate Stewart, Boyd Gilman, Stacy Dale, Randall Brown, Lauren Vollmer and Margaret Coit

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

THE BOARD OF TRUSTEES AT THE UNIVERSITY OF ILLINOIS, CHICAGO

The Board of Trustees at The University of Illinois, Chicago (UIC) received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Coordination of Health Care for Complex Kids (CHECK) program. The goal of the program was to improve care coordination for children and young adults who have chronic medical conditions. The target population consisted of Medicaid enrollees from birth to age 25 who were residents of Cook County, Illinois, and had one of four conditions: asthma, diabetes, sickle cell disease, or prematurity. The CHECK program launched in December 2014, four months after award. The intervention period funded by HCIA R2 ended in August 2018. Table 1 summarizes the key characteristics of the program.

The awardee hypothesized that addressing social determinants of health through better access to and coordination with social services, in addition to coordinating primary, specialty, and mental health care, would result in better health and health-related outcomes. Community health workers coordinated these nonmedical, medical, and mental health services, with oversight and guidance from care coordinators. The CHECK program also implemented software and consumer-facing technology to support its care coordination activities. The goals of the program were to (1) increase the number of participants and caregivers who were actively engaged in their own care; (2) improve participants' health and quality of life, including greater school attendance; and (3) reduce the total cost of care for participants by reducing hospitalizations and emergency department (ED) visits.

Important issues for understanding the evaluation

- The CHECK program aimed to improve health and quality of life, and reduce Medicaid costs for children and young adults with asthma, diabetes, sickle cell disease, or prematurity through improved coordination of medical, nonmedical, and mental health services.
- This impact analysis is based on a randomized controlled trial (RCT) started in the first half of the second program year (April 2016–March 2018), after early implementation challenges were resolved.
- This impact analysis is based on 6,259 Medicaid enrollees in one Medicaid managed care plan in Illinois, of whom 3,131 were assigned to treatment and 3,128 to control in April 2016. Only one-quarter of the treatment group members (N = 789) were actively engaged. This made it more difficult to detect program effects.

This impact analysis is limited to Medicaid recipients who participated in a randomized controlled trial (RCT) arm of the CHECK program and to outcomes that can be measured with Medicaid data (see Appendix A for details). All eligible enrollees were assigned to either treatment or control groups at the same time in April 2016—coinciding with the first half of the second year of the program. Enrollees assigned to the treatment group received the intervention. Control group enrollees received usual care.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The Board of Trustees of the UIC implemented the CHECK program to improve care coordination for children and young adults ages birth to 25 on Medicaid who had complex medical conditions in Cook County, Illinois.
Major innovation	The CHECK program deployed community health workers from the same neighborhoods as the children and families it served. These staff built relationships with children and their families, identified their needs for health and social service care coordination, and connected participants to information and services to address these needs. The program used health technology to share resources, provide communication support, and enable remote access to mental health services.
Program components	• Care coordination to address medical, mental, and social support needs • Enhanced mental health services and referrals • Health technology (videoconferencing platform, online self-education portal, two-way text messaging, and care coordination software) to support care delivery components
Target population	The program sought to engage children and young adults (ages 25 and younger) with chronic medical conditions. Participants met the following criteria: • Asthma (87 percent), diabetes (6 percent), premature birth (5 percent), or sickle cell disease (1 percent) • Enrolled either in a Medicaid managed care organization under contract to the CHECK program or in Medicaid fee-for-service
Total enrollment	The program enrolled 18,028 children and young adults (103 percent of original enrollment goal), including 3,131 who were assigned to the RCT treatment group..
Randomized enrollees	The awardee assigned 3,131 enrollees to the treatment group and 3,128 to the control group in April 2016. The awardee classified 4 percent as high risk, 22 percent as medium risk, and 74 percent as low risk. All randomized enrollees were in the Medicaid managed care plan at the time of enrollment.
Level of engagement	Of the 3,131 participants assigned to the treatment group, 789 (25 percent) had generated a care plan, 35 (1 percent) had an assessment but had not generated a care plan, and 2,307 (74 percent) did not have an assessment or a care plan. The analysis included all participants, regardless of level of engagement.
Theory of change or theory of action	Better access to social services and to primary, specialty, and mental health care will result in better health and social outcomes, including fewer hospitalizations and ED visits, and lower costs. Enhanced care coordination will address the medical and nonmedical needs of children and young adults with chronic medical conditions, including mental health care needs. The effectiveness of care coordination teams will be enhanced by using community health workers to coordinate medical and nonmedical services, working with care coordinators who led each care team, will enhance the effectiveness of care coordination teams. The CHECK program's new software and consumer-facing technology will enhance the effectiveness of its care coordination activities.
Award amount	$19,581,403
Effective launch date	• The program began operating December 2014. • The impact analysis was based on the two years of the RCT, which began in April 2016, and excluded almost 15,000participants who were not part of the RCT population.
Program settings	Community health workers engaged participants and families over the phone, as well as in participants' homes or in other community settings, including school- and community-based health centers. Staff on the mental health promotion team provided direct promotional and early intervention services to participants by phone and in the CHECK offices. Care coordination and mental health services were also provided through health technology (for example, short message service (text) platform and videoconference).
Market area	Urban, Cook County, Illinois

Table 1 (*continued*)

Program characteristics	Description
Target outcomes	• Increase the number of participants and families actively engaged in their own care • Improve participants' health and quality of life, including improving school attendance • Reduce total cost of care for the patient population through reduction in hospitalizations and ED visits
Payment model	A per beneficiary per month care coordination fee to be paid by Medicaid managed care organizations after end of award (fee-for-service Medicaid recipients would not be eligible)
Sustainability plans	At the end of the award, UIC planned to use internal funds to maintain CHECK services in the short term, while negotiating agreements with Medicaid managed care plans on its payment model for longer term support.

CHECK = Coordination of Health Care for Complex Kids; ED = emergency department; RCT = randomized controlled trial; UIC = University of Illinois, Chicago.

Table 2 describes key features of the CHECK evaluation. The impact analysis conducted as part of this evaluation included 6,259 children with at least one of four targeted conditions – asthma, diabetes, sickle cell disease, or prematurity – enrolled in one Medicaid managed care plan in Chicago, Illinois, 3,131 of which were randomized to the treatment group and 3,128 to the control group.

Table 2. Key features of the program evaluation

Features	Description
Evaluation design	The analysis relied on a randomized control group design.
Intervention group for evaluation	The treatment group included 3,131 Medicaid-covered children with at least one of four targeted conditions – asthma, diabetes, prematurity, or sickle cell disease.
Control group	The impact analysis compared outcomes among treatment group participants to those of a control group of 3,128 Medicaid-covered children who met the same study inclusion criteria.
Limitations	Because this study only evaluated outcomes measurable in claims and encounter records, it does not evaluate other important outcomes, such as improvements in quality of life and school attendance.

.

PROGRAM DESIGN AND ADAPTATION

The CHECK service delivery model had three key components: (1) care coordination, (2) enhanced mental health services, and (3) health technology.[1] However, the program enhanced the model in several ways over the course of the demonstration, which could have led to different impacts.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the CHECK program; the report is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Care coordination

Community health workers conducted an initial assessment of participants to identify how connections to social service agencies might address their nonhealth needs and to identify their physical and mental health needs. The community health workers then worked with participants and their families to connect them to relevant social service agencies and coordinate physical and mental health services, including the enhanced mental health services provided by the program. For each participant, the community health workers documented all coordination-related activities and information in a care plan.

Community health workers interacted with participants via telephone calls and in-person visits. The frequency of contact with the program varied based on participants' needs. Some participants required both in-person contacts and ongoing telephonic support. The community health workers deemed other participants to need only occasional telephone calls or use of the health technology component of the program. For in-person meetings, community health workers met participants at their homes, at social service agencies, in the hospital, at community- and school-based health centers, and at other local sites that were convenient for the participants. To address problems with low engagement rates, the awardee introduced program changes during the program. In the second year of the program, the community health workers attended clinical rounds in the pediatric units of the university hospital to reach out to participants who had been admitted. By the third year, the awardee had created a separate, dedicated team of community health workers focused solely on going into the community to find participants who were difficult to reach or had fallen out of touch with the CHECK program.

Enhanced mental health services

Program staff delivered a range of mental health services corresponding to needs identified at assessment. Services evolved from educating participants during the first year (for example, by using informational DVDs on childhood behavior distributed to new parents) to conducting regular mental health assessments, consulting with care coordination staff and participants' health care providers, and providing services and referrals in the second year and thereafter. In addition, in the third year of the program the awardee began offering expanded services such as one-on-one or online education in stress management and self-care (all participants), consultation (for one-time requests from community health workers or participants' health care providers), and referrals to longer-term mental health treatment or services.

Health technology

The program had a team of four information technology specialists working with a number of contracted technology vendors on care coordination software and other health technology to support the CHECK program. These program staff worked with a technology vendor to customize an existing care coordination software product to the needs of the CHECK program. The care coordination software enabled community health workers to document assessments, input care plans, and track ongoing contacts with participants and their families. The software

included an integrated repository of social service resources that community health workers used to provide and track referrals. In addition, CHECK offered health technology tools, such as education materials for patients, accessible online through the program's patient portal and a two-way text messaging platform, as well as videoconferencing technology to support participants' engagement in the program. Despite efforts to improve the software, program leaders noted that the software ultimately never functioned as intended and, thus, would likely have little effect on outcomes.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Awardee staff reported experiencing challenges engaging and activating patients throughout the program. In particular, staff reported that participants' contact information was often out of date or incomplete, making it difficult to find and communicate with them. Program staff reported that some participants' families might have been wary of the program's outreach, particularly if the family was concerned about government intervention in their lives (for example, a family might have feared that CHECK staff were from an immigration or child welfare agency). Further, the intervention focused on children and young adults, which meant the program had to engage their caregivers. Staff reported that caregivers might have had their own medical or behavioral issues which, despite the program's efforts, could have affected their willingness and ability to engage in CHECK.

Program leaders and staff also reported significant implementation delays during the first year of the program (before the RCT began). These included heavy staff workloads, lack of clarity about staff roles and responsibilities, lack of support and supervision for community health workers, and technical problems with the program's care coordination software that burdened staff with additional documentation tasks. These delays made it difficult for staff to deliver program services. In the second year, the program underwent a significant change in leadership within the care coordination and mental health promotion teams. The new leaders actively addressed many of the issues plaguing the program. They also restructured the program's service delivery teams, staff oversight, and training, and redesigned program protocols. Awardee staff reported that these implementation changes led to

Implications of program implementation for detecting impacts

- The low level of participant engagement (25 percent of enrollees) limited the program's effectiveness and made it more difficult to detect impacts.
- Delaying the evaluation period until April 2016 (nearly 20 months after award) after many of the early intervention delivery problems were resolved increased the likelihood that the program was effective.
- If participants received more and more timely services as the RCT progressed, impacts might be greater in the second year of the RCT, April 2017–March 2018 (that is, in the third year of program operations).

improvements in outreach and delivery of intervention services over time. By the beginning of the third year of the program (during the RCT), staff reported delivering ongoing care coordination and mental health services successfully.

ESTIMATING PROGRAM IMPACTS

Study sample

In April 2016, UIC and one of its Medicaid managed care partners, Harmony, agreed to randomly assign all eligible children and young adults into treatment and control groups. Harmony identified all of its members who were eligible for CHECK as of April 2016 based on their claims history and who had not yet been enrolled in CHECK. Using the full list of eligible members from Harmony, the independent evaluation team randomly assigned at the same time 3,131 eligible members of the managed care plan to the treatment group and 3,128 members to the control group. To ensure balance across the two groups, the children and young adults were first stratified by risk tier. The awardee defined medium-risk enrollees as those who had one to three ED visits or one inpatient hospitalization in the year before randomization. The awardee defined high-risk patients as those who had more than three ED visits or more than one inpatient hospitalization in the year before randomization. Before enrollment, the awardee classified the remaining enrollees as low risk. Within each tier, children and young adults were assigned to treatment and control groups, accounting for the distribution of the target conditions, gender, age category, and ZIP code of residence in each tier. UIC attempted to enroll into the program all children and young adults who were assigned to the treatment group.

Enrolling, engaging, and activating participants

The awardee classified participants as either enrolled, engaged, or activated. At the start of the RCT, community health workers enrolled participants in the study by sending an invitation letter to them or their caregivers. The community health workers subsequently engaged enrolled participants by starting an assessment via phone or in person. Finally, engaged participants were deemed to be activated when the community health worker had generated a care plan for them. More than half of the treatment group had some contact with the community health workers, but only one-quarter (789) completed an assessment and had a care plan in place. Another 1 percent (35) were engaged but not activated, usually due to loss of contact with patients and their families. The awardee considered the remaining treatment group members (2,307) enrolled only, either because the program could not reach them or they refused to engage.

The engagement and activation rates in the combined medium- and high-risk subgroup were about the same as in the full treatment group. Across the full sample and the combined medium- and high-risk subgroup, participants who were in younger age categories (younger than 19) and from ZIP codes with lower median incomes and larger Hispanic populations were more likely to become engaged or activated (data not shown).

Most contact with a community health worker occurred during the first year of the RCT and was primarily by telephone. The mean number of contacts among activated participants, both in the full sample and the combined medium- and high-risk subgroup, was about 3.5 during the first year of the RCT and about 1.0 during the second year. The impact analysis includes all treatment group members, including those who were only enrolled and not engaged or activated, to ensure the estimates are unbiased.

Characteristics of treatment and control group members

A comparison of treatment and control group characteristics at baseline confirmed that the two groups were well balanced (Table 3). The combined subgroup of medium- and high-risk participants in the RCT was also well balanced between the treatment and control groups. Appendix B provides the full balance results measured during the 12 months before enrollment in the RCT.

The average age of treatment and control group members during the baseline year was 11 years. Asthma was by far the most common target condition among enrollees, accounting for 86 percent of the full RCT group. High-risk enrollees comprised only 4 percent of the RCT group and medium-risk enrollees accounted for 22 percent. The awardee classified the remaining 74 percent as low risk. In the following discussion, the *higher-risk groups* refers to the combined group of medium- and high-risk enrollees. The average Chronic Illness and Disability Payment System (CDPS) score was 2.7 for the treatment group and 2.5 for the control group, indicating that their expected Medicaid spending was at least 2.5 times higher than average spending for children on Medicaid nationally. The average CDPS score was 3.6 and 3.3 for the treatment and control groups, respectively, in the higher-risk subgroups.

About 6 percent of the full RCT group had a hospitalization during the baseline year, and slightly less than 35 percent in both treatment and control groups had an outpatient ED visit. The rates were higher in the combined medium- and high-risk subgroups, in which 18 percent had a hospitalization and 72 percent had an outpatient ED visit. (Some beneficiaries in the higher-risk groups did not have a hospitalization or ED visit in the year before enrollment because the data used to define risk status differed from the data used to capture service use in the baseline period.) Average spending per beneficiary per month (PBPM) among the full RCT group was $183 for the treatment group, compared to $203 for the control group. The regression model includes the variable as a control to account for this difference and to improve the precision of the estimates. Average PBPM spending in the higher-risk subgroups was about twice as high as in the full RCT group.

Table 3. Baseline characteristics of treatment and control group members

Measure	Full group		Medium- and high-risk groups	
	Treatment (N = 3,131)	Control (N = 3,128)	Treatment (N = 821)	Control (N = 819)
Demographics				
Age at enrollment, years	11	11	11	11
Age group, %				
0–8 years	39	39	43	42
9–11 years	16	15	11	11
12–18 years	31	31	22	24
19–25 years	15	15	24	23
Male, %	53	53	50	49
Target conditions, %				
Asthma	86	87	84	84
Diabetes	6	6	6	6
Prematurity	5	5	6	6
Sickle cell disease	1	1	1	1
Risk tier, %				
High	4	4	17	17
Medium	22	22	83	83
Low	74	74	n.a.	n.a.
Health status, service use, and expenditures during the year before enrollment				
CDPS score[a]	2.7	2.5	3.6	3.3
Any hospitalizations, %	5	6	17	18
Any outpatient ED visits, %	34	33	73	71
Number of outpatient ED visits	0.6	0.6	1.6	1.5
Total Medicaid expenditures ($ PBPM)	$183	$203	$372	$391

Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2015, to March 31, 2016.

Notes: The baseline period covers the 12-month period from April 1, 2015, to March 31, 2016.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid.

None of the differences between treatment and control groups in any of the baseline characteristics was statistically different from zero at the 0.10 level, two-tailed test.

Appendix B presents the full balance results.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; n.a. = not applicable; PBPM = per beneficiary per month.

Analytic approach

The RCT study design ensures that impact estimates are unbiased. The study outcomes, described in Appendix A, were obtained from Medicaid claims and encounter records. The impact estimates were obtained from a regression of key outcomes on enrollees' characteristics, and cover the 24-month period from April 1, 2016, to March 31, 2018. Differences between the

separate estimates for the first and second years of the follow-up period reveal how program impacts vary with participants' length of exposure to the intervention or to changes in program implementation. Separate estimates were obtained for higher-risk participants to assess whether impacts were larger for these groups, as nearly all studies of care coordination interventions have found.[2] Appendix A contains a detailed description of the statistical model used to estimate the effects of the program. The program impacts for UIC were also estimated using a Bayesian approach, and presented in Appendix D.

IMPACT RESULTS

CHECK appeared to have favorable impacts *among the combined higher-risk subgroups* on total Medicaid spending, inpatient admissions, and outpatient ED visits, particularly in Year 1 of the RCT period (Table 4). Not all of the impact estimates were statistically significant, but the point estimates consistently indicated favorable findings. This was true even when data were truncated for outliers. For example, the findings suggest that CHECK reduced total Medicaid expenditures by about **19 percent** for the higher-risk groups in Year 1. When spending outliers in Year 1 were truncated, the impact estimate was still favorable, though it fell to about 6 percent and was not significant.[3] In Year 1 for the higher-risk groups, the probability of any hospitalization and any outpatient ED visit decreased significantly by **2.5** and **4.7 percentage points**, respectively. During that same period, the estimates for the number of hospitalizations and ED visits also declined for the higher-risk groups; however, they were not statistically significant, possibly because only one-fourth of the treatment group received the intervention, which substantially decreased the power to detect program effects of a given size on the treatment group patients who actually received the intervention.[4] *Among the full sample*, the CHECK program had no discernible impacts on Medicaid expenditures, hospitalizations, or ED visits over the 24-month period or separately by year.

[2] Brown, R.S., D. Peikes, G. Peterson, J. Schore, and C. Razafindrakoto. "Six Features of Medicare Coordinated Care Demonstration Programs that Cut Hospital Admissions of High-Risk Patients." *Health Affairs,* vol. 31, no. 1156, 2012. Counsell, S.R., C.M. Callahan, W. Tu, T.E. Stump, and G.W. Arling. "Cost Analysis of the Geriatric Resources for Assessment and Care of Elders Care Management Intervention." *Journal of the American Geriatrics Society,* vol. 57, no. 8, 2009, pp. 1420–1426.

[3] There were 60 expenditure outliers truncated at the 98th percentile (29 in the treatment group and 31 in the control group). The outliers were larger on average for the control group than for the treatment group ($2,550 versus $1,750, respectively).

[4] If CHECK reduced hospitalizations by 20 percent for the 25 percent of treatment group actually receiving the intervention, the effect measured over the full sample would be only 5 percent. The power to detect an effect this small in the sample of 1,640 higher-risk children is only 10 percent.

Table 4. Estimated percentage impact of CHECK on selected outcome measures

	Full group (N = 6,259)			Medium and high-risk groups (N = 1,640)		
	Months 1–24	Months 1–12	Months 13–24	Months 1–24	Months 1–12	Months 13–24
Expenditures (PBPM)						
Impact ($)	-$6.0	-$5.8	-$6.3	-$44**	-$42	-$47*
Percentage impact	-4.0%	-3.8%	-4.3%	-21%	-19%	-23%
p-value	0.66	0.74	0.74	0.05	0.10	0.08
Any hospitalization						
Impact (percentage point)	0.42	-0.27	0.79	-0.60	-2.5*	1.3
Percentage impact	6.0%	-6.5%	21%	-4.7%	-28%	20%
p-value	0.52	0.58	0.13	0.70	0.05	0.31
Any ED visit						
Impact (percentage point)	-0.91	-1.6	-0.49	-2.8	-4.7**	-2.7
Percentage impact	-1.8%	-4.7%	-1.5%	-4.5%	-9.7%	-6.1%
p-value	0.46	0.17	0.68	0.22	0.04	0.27
Number of hospitalizations, per 1,000 beneficiaries						
Impact (count)	4.0	-0.16	8.5	-17	-28	-5.3
Percentage impact	8.0%	< 1%	18%	-15%	-22%	-5.4%
p-value	0.50	0.99	0.27	0.32	0.25	0.81
Number of ED visits, per 1,000 beneficiaries						
Impact (count)	-3.5	-22	16	-56	-106	-3.6
Percentage impact	< 1%	-3.6%	2.8%	-5.7%	-10%	< 1%
p-value	0.87	0.40	0.58	0.31	0.12	0.96

Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2016, to March 31, 2018.

Notes: Impact estimates are based on the regression-adjusted difference between the randomized treatment and control group members. Two years of follow-up data were available for all sample members. Percentage impacts were then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents the full impact estimates. Appendix D shows the results from the Bayesian analysis.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

CHECK = Coordination of Health Care for Complex Kids; ED = emergency department; PBPM = per beneficiary per month.

The favorable findings among the higher-risk groups are consistent with the expectation that impacts are likely to be concentrated among sicker patients. Program leaders noted that the tiered intake system identifying medium- and high-risk patients based on their health status, psychosocial needs, and history of ED visits and hospitalizations was critical for identifying and responding to those individuals most in need of services. According to the awardee, these risk tiers sought to help care coordination staff target their efforts to participants who were most in need of program support. For example, although community health workers were expected to call low-risk participants every 90 days, they were expected to call or visit high-risk beneficiaries

every 30 days. In addition, to accommodate the extra time needed to serve higher-risk participants, program leaders hired additional community health workers and care coordinator supervisors, giving community health workers more time and support to follow up and build relationships with sicker participants. Community health workers serving higher-risk participants also had smaller caseloads (about 175 cases) than those serving low-risk beneficiaries (250 to 400 cases). The awardee made similar staffing and caseload changes with the mental health promotion team during the RCT.

Main findings from impact evaluation

- CHECK had no discernable impacts on outcomes for the full RCT sample.
- CHECK appears to have reduced the probability of any hospitalization and any outpatient ED visit among the higher-risk subgroups in Year 1 of the RCT period. CHECK also appeared to reduce total Medicaid expenditures among the higher-risk subgroups, but the magnitude and statistical significance of the impact estimates declined substantially when accounting for the effects of outliers.
- The favorable estimates for the higher-risk groups are consistent with previous studies that reported favorable findings were generally concentrated among sicker patients. Effects were concentrated in the first year after enrollment in the RCT, which was when the community health workers had more contacts with activated treatment group participants.

The favorable impacts during the first year of the RCT coincides with the period when the community health workers had most of their contact with activated treatment group members. This suggests that the program might have helped enrollees and families to avoid using these services in that first year. However, as time passed and contact decreased, utilization levels became similar to those in the control group. In the higher-risk subgroups, the mean hospitalization and ED visit rates for the control group declined substantially from the first to the second year (see Appendix C), increasing the difficulty of reducing service use and related expenditures in the second year.

Starting the trial after the awardee fully implemented the program enhanced CHECK's ability to demonstrate favorable impacts among the higher-risk participants. The RCT started in April 2016, more than 15 months after the program began. Thus, the RCT did not overlap with the first year of the intervention, when there was less infrastructure, fewer resources, and more confusion on the part of staff about their roles and responsibilities. During this early period, the program made several changes that improved implementation. Halfway through the first program year (before the start of the RCT), program leadership began restructuring the workload, processes, and structure of the care coordination teams. For example, program leaders developed and refined protocols for staff to prioritize their contact efforts and decide when to stop reaching out to potential participants. These protocols enabled staff to better manage their time between delivering services to engaged and activated patients and conducting outreach. CHECK program leaders also made several changes to existing practices to meet participants' observed needs, including creating a dedicated care coordination team focused solely on hard-to-reach participants, and shifting some care coordination staff to evening or weekend shifts when they might be more likely to reach participants.

CONCLUSION

The CHECK program demonstrated modest impacts on hospitalizations, ED visits, and expenditures for the subgroups of higher-risk participants enrolled in the RCT. These limited effects are likely due to the fact that the program engaged or activated only one-quarter of higher-risk participants in the treatment group. Assuming that program effects were confined solely to those who actually received intervention services, the impacts on these engaged and activated enrollees would have been four times larger than the estimated impact for the full study group. For this reason, the favorable impacts should be interpreted with some caution. The generally inconclusive findings for total Medicaid spending and number of hospitalizations and number of ED visits for the higher-risk subgroups might also be due to the relatively low engagement rate. Specifically, the relatively small share of engaged or activated treatment group participants substantially reduces the power to detect impacts of meaningful size.

The low engagement rate suggests that the estimated effects might have been substantially larger if the program had actively engaged more than one-fourth of the treatment subgroups' members. However, the extent to which this might be true is unknown. Programs do not affect all program participants equally, so perhaps the CHECK program successfully reached all or most of those enrollees and families who were willing to participate and would benefit from the program. In that case, the current impact estimates reflect the likely impacts over the full target population. Alternatively, there could be other participants and families in the treatment group for whom the program could have helped improve outcomes if the awardee had been able to engage them. In this case, the impact estimates might understate the potential impacts of the intervention.

Limitations of evaluation

Randomized controlled trials are often considered the gold-standard for estimating program impacts. However, this study was limited to assessing impacts on outcomes measurable in claims and encounter records. To the extent that the intervention had favorable impacts on other, non-claims-based outcomes, such as quality of life and school attendance, this evaluation could not identify those types of impacts. Similarly, the study could not measure any potential spillover impacts on siblings of treatment group children or on parents' quality of life and ability to work.

PROGRAM SUSTAINABILITY

UIC reported that after the end of the award in August 2018 it intended to use internal funding to maintain CHECK services in the short term, while waiting to reach agreements with payers on its payment model for longer-term support. However, UIC's negotiations with the managed care plans remained on hold at the end of the award because the state Medicaid agency delayed implementing its managed care contracts. During this delay, UIC continued preparing for the negotiations, promoting the program to the managed care plans and hiring a third party to facilitate the negotiations.

UIC's proposed payment model

UIC proposed to fund CHECK services through a PBPM fee that Medicaid managed care organizations would pay to the awardee for each enrolled beneficiary. With support from an actuarial consultant, the awardee estimated the PBPM amount would depend on the number of children and young adults enrolled, and range from $23 to $55, more than triple the estimated savings in total Medicaid expenditures of $6 PBPM over the full sample. The awardee did not propose any adjustments to the PBPM fee based on enrollees' medical complexity, quality measures, or spending benchmarks.

The primary payment model proposed by the awardee was a PBPM fee that a Medicaid managed care organization would pay to the program for each enrolled beneficiary. The fee would cover the same services implemented under the award for the same target populations. With support from an actuarial consultant, the awardee estimated the payment rate at $23 PBPM when the program was fully operational in Year 3. Based on the awardee's actuarial estimates, the estimated savings in Medicaid expenditures over the full sample would not be enough to cover the costs of the program, whereas the estimated savings for higher-risk enrollees of $44 PBPM would be more than enough to cover the average costs per beneficiary. However, if the program were limited to high-risk beneficiaries, the average costs per beneficiary would likely be much higher than $23 PBPM because a program serving only high-risk beneficiaries would need more care coordinators. Thus, the program's sustainability depends on the size of the higher-risk populations that CHECK would serve, whether program could sustain the effects over the payment period, and whether the program could engage a higher proportion of eligible children and young adults and their caregivers. States serving Medicaid recipients in fee-for-service might also want to consider testing the program, given the potential for net savings demonstrated in this RCT. However, it is not clear if a fee for service environment, where financial incentives are much different, can generate similar effects.

This page has been left blank for double-sided copying.

Appendix A

Definition of outcomes measures and description of modeling strategy and control variables

This page has been left blank for double-sided copying.

1. Outcomes measures

The analysis file was constructed using data provided by the University of Illinois, Chicago as well as publicly available data. The regression model included control variables constructed from the randomization file (demographic characteristics, target conditions, secondary conditions, and risk tier); eligibility file (monthly enrollment in Medicaid); and Medicaid claims data (pre-enrollment health care use and spending). The model also included ZIP code-level data from the publicly available American Community Survey.

The outcomes for this evaluation included three core measures (total Medicaid spending, number of hospitalizations, and number of outpatient emergency department [ED] visits) and two awardee-specific measures (number of asthma-related admissions and outpatient ED visits). The analysis used Medicaid claims data to construct all of the outcomes. All enrollees in the randomized controlled trial were enrolled in managed care at the time of enrollment in the Coordination of Health Care for Complex Kids (CHECK) program and most remained in managed care throughout the baseline and intervention periods. However, about 20 percent of participants in both the treatment and control groups were enrolled in fee-for-service (FFS) for one or more months during the evaluation. Those who switched to FFS after enrollment remain in the impact evaluation. Medicaid spending in this analysis reflects the sum of all FFS spending from Medicaid claims and the payment variable on the managed care encounter records for managed care enrollees. The latter reflects what the FFS payment would have been for that service, not necessarily what the managed care organization actually paid.

2. Modeling strategy and control variables

In the randomized controlled trial approach, the awardee randomly assigned eligible plan members into either a treatment group or a control group. Program effects were estimated using regression models of the following form:

$$(1) \qquad Y_{it} = \alpha + \theta_t Treatment_i + B' X_i + \gamma' C_i + \varepsilon_{it}$$

where Y_{it} is the outcome of individual *i* in period *t* (for example, total monthly Medicaid expenditures during the *t*-th time period since he or she enrolled); α is a constant term; $Treatment_i$ is an indicator for whether the individual is assigned to the group that received program services; X_i are beneficiary characteristics including age, gender, risk group (high, medium, or low), Chronic Illness and Disability Payment System score, target condition, secondary conditions, and other pre-enrollment characteristics, including baseline values of outcome measures; and C_i are community characteristics that can affect outcomes (including percentage of the ZIP code residents who are white, Black, Hispanic, or other race or ethnicity). ε_{it} is a random disturbance term.

Equation (1) was estimated separately for each period in order to produce an estimate of all parameters for each yearly period $T = \{1,2,\}$ following the start of the randomization period. The key parameter of interest is θ_t, which measures the impact of the program in participants' *t*-th period after enrolling. Thus, the model specification allowed an assessment of how program impacts varied with enrollees' length of exposure to the program. This report presents results for the two annual follow-up periods. The model was also used to estimate effects over the full 24-month period by constructing outcomes for this longer period. The analysis produced estimated impacts for the subgroups of higher-risk patients by adding interaction terms to the model.

Appendix B

Results from balance assessment of treatment and control group

This page has been left blank for double-sided copying.

Table B.1. Baseline characteristics of all beneficiaries in treatment and control groups for UIC

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	11 (0.11)	11 (0.11)	-0.04 (0.16)	< +/-1	-0.01	0.80	< 0.01
Age: 0 to 8, %	39 (0.87)	39 (0.87)	0.19 (1.2)	< +/-1	0.00	0.88	< 0.01
Age: 9 to 11, %	16 (0.65)	15 (0.64)	0.30 (0.93)	2.0	0.01	0.74	< 0.01
Age: 12 to 18, %	31 (0.83)	31 (0.83)	-0.25 (1.2)	< +/-1	-0.01	0.83	< 0.01
Age: 19 to 25, %	15 (0.63)	15 (0.64)	-0.24 (0.91)	-1.6	-0.01	0.79	< 0.01
Male, %	53 (0.89)	53 (0.89)	0.11 (1.2)	< +/-1	0.00	0.93	< 0.01
Risk tier, %							
High	4.4 (0.37)	4.4 (0.37)	0.03 (0.49)	< +/-1	0.00	0.96	< 0.01
Medium	22 (0.74)	22 (0.74)	0.01 (1.1)	< +/-1	0.00	0.99	< 0.01
Low	74 (0.79)	74 (0.79)	-0.04 (1.2)	< +/-1	0.00	0.97	< 0.01
Health status and diagnoses							
CDPS score[a]	2.7 (0.07)	2.5 (0.05)	0.15 (0.09)	5.7	0.05	0.07	< 0.01
Asthma, %	86 (0.62)	87 (0.61)	-0.66 (0.88)	< +/-1	-0.02	0.45	< 0.01
Diabetes, %	5.7 (0.41)	5.7 (0.41)	-0.01 (0.58)	< +/-1	0.00	0.99	< 0.01
Prematurity, %	4.7 (0.38)	4.6 (0.38)	0.03 (0.54)	< +/-1	0.00	0.96	< 0.01
Sickle cell disease, %	0.67 (0.15)	0.80 (0.16)	-0.13 (0.21)	-19	-0.02	0.55	< 0.01
Brain injury, %	0.45 (0.12)	0.22 (0.08)	0.22 (0.15)	50	0.04	0.13	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Seizure, %	4.2 (0.36)	4.1 (0.35)	0.09 (0.50)	2.2	0.00	0.86	< 0.01
Medicaid expenditures (PBPM)							
Total expenditures	187 (22)	206 (27)	-20 (33)	-11	-0.01	0.57	< 0.01
Total inpatient expenditures	74 (21)	83 (25)	-9.8 (31)	-13	-0.01	0.76	< 0.01
Total ED or observation expenditures	9.2 (0.50)	9.0 (0.43)	0.27 (0.66)	3.0	0.01	0.68	< 0.01
Total prescription drug expenditures	30 (1.8)	34 (2.5)	-4.2 (3.1)	-14	-0.03	0.17	< 0.01
Total all other Medicaid expenditures	74 (2.6)	80 (4.4)	-6.0 (5.0)	-8.1	-0.03	0.25	< 0.01
Service utilization							
Any hospitalization, %	5.5 (0.41)	5.8 (0.42)	-0.32 (0.58)	-5.9	-0.01	0.58	< 0.01
Total hospitalizations	0.07 (0.01)	0.08 (0.01)	-0.01 (0.01)	-15	-0.03	0.28	< 0.01
Any ED or observation visit, %	33 (0.84)	33 (0.84)	0.51 (1.2)	1.5	0.01	0.67	< 0.01
Total ED or observation visits	0.62 (0.02)	0.59 (0.02)	0.03 (0.03)	4.4	0.02	0.38	< 0.01
Any asthma-related hospitalization, %[b]	2.2 (0.26)	2.4 (0.27)	-0.23 (0.38)	-10	-0.02	0.55	< 0.01
Total asthma-related hospitalizations[b]	0.03 (0.00)	0.03 (0.00)	-0.01 (0.01)	-22	-0.03	0.28	< 0.01
Any ED or observation visits for asthma, %[b]	11 (0.56)	11 (0.56)	0.18 (0.79)	1.6	0.01	0.82	< 0.01
Total ED or observation visits for asthma[b]	0.16 (0.01)	0.15 (0.01)	0.01 (0.02)	7.9	0.02	0.39	< 0.01
Zip code-level factors[c]							
Median income	45,810 (305)	45,812 (311)	-2.5 (430)	< +/-1	0.00	1.00	< 0.01
Percent White	26 (0.44)	25 (0.44)	0.16 (0.61)	< +/-1	0.01	0.80	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Percent Black	38 (0.62)	38 (0.61)	0.16 (0.85)	< +/-1	0.00	0.85	< 0.01
Percent Hispanic	32 (0.47)	33 (0.48)	-0.46 (0.65)	-1.4	-0.02	0.49	< 0.01
Percent other	4.5 (0.09)	4.4 (0.09)	0.14 (0.12)	3.0	0.03	0.28	< 0.01
Percent poverty	23 (0.19)	23 (0.19)	-0.10 (0.27)	< +/-1	-0.01	0.70	< 0.01
Percent high school degree or higher	77 (0.22)	77 (0.22)	0.21 (0.30)	< +/-1	0.02	0.49	< 0.01
Percent college degree or higher	20 (0.22)	20 (0.23)	0.05 (0.31)	< +/-1	0.00	0.89	< 0.01
Percent unemployed	15 (0.11)	15 (0.11)	0.00 (0.15)	< +/-1	0.00	0.98	< 0.01
Percent U.S. citizens	89 (0.17)	89 (0.17)	0.11 (0.23)	< +/-1	0.01	0.63	< 0.01
Percent uninsured	17 (0.11)	18 (0.11)	-0.10 (0.15)	< +/-1	-0.02	0.53	< 0.01
Number of beneficiaries	3,131	3,128					
Omnibus test				Chi-squared statistic 27.08	Degrees of freedom 40.00	P-value 0.94	

Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2015, to March 31, 2016.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

[b] Asthma-related hospitalizations and ED visits were identified based on all available diagnosis codes on the inpatient and ED claims.

[c] Zip code-level characteristics were obtained by merging beneficiaries' zip code of residence to the publicly available American Community Survey 5-year zip code estimates file (2011-2015).

CDPS = chronic illness and disability payment system; ED = emergency department; PBPM = per beneficiary per month; SE = standard error; UIC = University of Illinois at Chicago.

Table B.2. Baseline characteristics of medium- and high-risk beneficiaries in treatment and control groups for UIC

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	11 (0.27)	11 (0.26)	0.01 (0.38)	< +/-1	0.00	0.99	< 0.01
Age: 0 to 8, %	43 (1.7)	42 (1.7)	0.99 (2.5)	2.3	0.02	0.68	< 0.01
Age: 9 to 11, %	11 (1.1)	11 (1.1)	-0.15 (1.5)	-1.4	0.00	0.92	< 0.01
Age: 12 to 18, %	22 (1.4)	24 (1.5)	-1.6 (2.0)	-7.4	-0.04	0.43	< 0.01
Age: 19 to 25, %	24 (1.5)	23 (1.5)	0.80 (2.1)	3.3	0.02	0.70	< 0.01
Male, %	49 (1.7)	49 (1.7)	0.37 (2.5)	< +/-1	0.01	0.88	< 0.01
Risk tier, %							
High	17 (1.3)	17 (1.3)	0.08 (1.8)	< +/-1	0.00	0.96	< 0.01
Medium	83 (1.3)	83 (1.3)	-0.08 (1.8)	< +/-1	0.00	0.96	< 0.01
Health status and diagnoses							
CDPS score	3.6 (0.16)	3.3 (0.11)	0.35 (0.19)	9.7	0.09	0.07	< 0.01
Asthma, %	84 (1.3)	84 (1.3)	0.04 (1.8)	< +/-1	0.00	0.98	< 0.01
Diabetes, %	6.0 (0.83)	6.1 (0.84)	-0.14 (1.2)	-2.3	-0.01	0.91	< 0.01
Prematurity, %	6.0 (0.83)	5.9 (0.82)	0.11 (1.2)	1.8	0.00	0.93	< 0.01
Sickle cell disease, %	0.73 (0.30)	1.3 (0.40)	-0.61 (0.51)	-84	-0.06	0.22	< 0.01
Brain injury, %	0.85 (0.32)	0.12 (0.12)	0.73 (0.35)	86	0.10	0.03	< 0.01
Seizure, %	5.0 (0.76)	6.1 (0.84)	-1.1 (1.1)	-22	-0.05	0.33	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Medicaid expenditures (PBPM)							
Total expenditures	379 (52)	400 (57)	-21 (78)	-5.5	-0.01	0.79	< 0.01
Total inpatient expenditures	195 (48)	200 (50)	-4.2 (70)	-2.1	0.00	0.95	< 0.01
Total ED or observation expenditures	26 (1.7)	21 (1.0)	4.3 (1.9)	17	0.11	0.03	< 0.01
Total prescription drug expenditures	39 (4.5)	52 (8.1)	-12 (9.6)	-31	-0.07	0.18	0.02
Total all other Medicaid expenditures	118 (7.9)	127 (10)	-8.6 (13)	-7.3	-0.03	0.50	< 0.01
Service utilization							
Any hospitalization, %	17 (1.3)	18 (1.4)	-0.90 (1.9)	-5.2	-0.02	0.64	< 0.01
Total hospitalizations	0.24 (0.02)	0.27 (0.03)	-0.04 (0.03)	-16	-0.05	0.29	< 0.01
Any ED or observation visit, %	72 (1.6)	69 (1.6)	2.6 (2.3)	3.7	0.06	0.24	< 0.01
Total ED or observation visits	1.6 (0.07)	1.5 (0.06)	0.13 (0.09)	8.1	0.07	0.15	< 0.01
Any asthma-related hospitalization, %[a]	7.1 (0.89)	7.7 (0.93)	-0.63 (1.3)	-8.9	-0.02	0.63	< 0.01
Total asthma-related hospitalizations[a]	0.08 (0.01)	0.10 (0.01)	-0.02 (0.02)	-23	-0.05	0.29	< 0.01
Any ED or observation visits for asthma, %[a]	27 (1.6)	26 (1.5)	1.2 (2.3)	4.2	0.03	0.60	< 0.01
Total ED or observation visits for asthma[a]	0.45 (0.04)	0.39 (0.03)	0.05 (0.05)	12	0.05	0.29	< 0.01
Zip code-level factors[b]							
Median income	48,069 (649)	47,765 (654)	304 (919)	< +/-1	0.02	0.74	< 0.01
Percent White	30 (0.94)	29 (0.94)	0.87 (1.3)	2.9	0.03	0.52	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Percent Black	35 (1.2)	34 (1.2)	0.90 (1.7)	2.6	0.03	0.59	< 0.01
Percent Hispanic	30 (0.90)	32 (0.92)	-1.9 (1.3)	-6.4	-0.07	0.13	< 0.01
Percent other	5.0 (0.18)	4.8 (0.17)	0.16 (0.25)	3.3	0.03	0.51	< 0.01
Percent poverty	22 (0.38)	22 (0.38)	-0.36 (0.55)	-1.6	-0.03	0.50	< 0.01
Percent high school degree or higher	78 (0.44)	77 (0.46)	0.93 (0.65)	1.2	0.07	0.14	< 0.01
Percent college degree or higher	21 (0.45)	20 (0.47)	0.31 (0.63)	1.5	0.02	0.64	< 0.01
Percent unemployed	14 (0.23)	14 (0.22)	0.04 (0.32)	< +/-1	0.01	0.89	< 0.01
Percent U.S. citizens	89 (0.32)	89 (0.33)	0.73 (0.47)	< +/-1	0.08	0.11	< 0.01
Percent uninsured	17 (0.23)	17 (0.23)	-0.51 (0.32)	-3.1	-0.08	0.12	< 0.01
Number of beneficiaries	821	819					
Omnibus test				Chi-squared statistic 37.96	Degrees of freedom 39.00	P-value 0.52	

Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2015, to March 31, 2016.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable.

[a] Asthma-related hospitalizations and ED visits were identified based on all available diagnosis codes on the inpatient and ED claims.

[b] Zip code-level characteristics were obtained by merging beneficiaries' zip code of residence to the publicly available American Community Survey 5-year zip code estimates file (2011-2015).

CDPS = chronic illness and disability payment system; ED = emergency department; SE = standard error; UIC = University of Illinois at Chicago.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Table C.1. Estimated impact of the UIC intervention on Medicaid expenditures and utilization measures during a 24-month follow-up period

	All randomized beneficiaries					Medium- and high-risk beneficiaries only				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)										
Year 1	146	151	-5.8 (17)	-3.8%	0.74	184	225	-42 (26)	-19%	0.10
Year 2	140	146	-6.3 (19)	-4.3%	0.74	159	206	-47* (26)	-23%	0.08
Cumulative	143	149	-6.0 (14)	-4.0%	0.66	172	216	-44** (23)	-21%	0.05
Acute inpatient expenditures ($ PBPM)										
Year 1	38	36	1.2 (15)	3.4%	0.94	37	66	-30* (17)	-45%	0.08
Year 2	36	26	11 (15)	42%	0.48	26	41	-14 (12)	-35%	0.22
Cumulative	37	31	5.9 (11)	19%	0.59	32	54	-22** (11)	-41%	0.04
Outpatient ED expenditures ($ PBPM)										
Year 1	8.8	8.9	-0.07 (0.61)	< 1%	0.91	16	16	-0.24 (1.8)	-1.5%	0.89
Year 2	9.9	8.5	1.4** (0.65)	17%	0.03	16	13	3.0* (1.6)	23%	0.06
Cumulative	9.3	8.7	0.65 (0.49)	7.4%	0.19	16	14	1.3 (1.3)	9.3%	0.32
Pharmacy expenditures ($ PBPM)										
Year 1	31	34	-2.5 (3.6)	-7.4%	0.49	42	51	-8.4 (11)	-17%	0.42
Year 2	31	43	-12 (9.0)	-27%	0.20	45	56	-11 (16)	-20%	0.48
Cumulative	31	38	-6.9 (5.2)	-18%	0.18	43	53	-9.8 (11)	-18%	0.38
All other Medicaid expenditures ($ PBPM)										
Year 1	68	73	-4.5 (4.4)	-6.1%	0.31	89	92	-3.6 (9.7)	-3.9%	0.71
Year 2	63	69	-6.8 (4.7)	-9.9%	0.15	72	96	-24* (13)	-25%	0.07
Cumulative	65	71	-5.6 (3.8)	-7.9%	0.14	81	94	-14 (10)	-15%	0.18
Hospital stays, per 1,000 beneficiaries										
Year 1	54	54	-0.16 (8.4)	< 1%	0.99	98	127	-28 (25)	-22%	0.25
Year 2	56	47	8.5 (7.7)	18%	0.27	92	97	-5.3 (22)	-5.4%	0.81
Cumulative	55	51	4.0 (6.0)	8.0%	0.50	95	112	-17 (17)	-15%	0.32

Table C.1 (*continued*)

	All randomized beneficiaries					Medium- and high-risk beneficiaries only				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays with an asthma diagnosis, per 1,000 beneficiaries[b]										
Year 1	19	23	-4.5 (4.7)	-19%	0.34	34	52	-18 (14)	-36%	0.19
Year 2	24	17	6.9 (4.5)	41%	0.12	27	29	-2.1 (9.6)	-7.1%	0.83
Cumulative	21	20	1.0 (3.3)	5.1%	0.76	31	41	-10 (8.7)	-25%	0.23
ED visits, per 1,000 beneficiaries										
Year 1	574	596	-22 (26)	-3.6%	0.40	952	1,058	-106 (69)	-10%	0.12
Year 2	586	571	16 (28)	2.8%	0.58	904	908	-3.6 (69)	< -1%	0.96
Cumulative	580	583	-3.5 (22)	< 1%	0.87	929	984	-56 (55)	-5.7%	0.31
ED visits with an asthma diagnosis, per 1,000 beneficiaries[b]										
Year 1	141	141	0.65 (13)	< 1%	0.96	231	244	-12 (34)	-5.1%	0.71
Year 2	146	135	11 (13)	8.2%	0.40	228	225	2.9 (32)	1.3%	0.93
Cumulative	144	138	5.7 (10)	4.2%	0.58	230	235	-4.9 (26)	-2.1%	0.85
Percentage of beneficiaries with any hospital admission in a time period										
Year 1	3.9	4.2	-0.27 (0.49)	-6.5%	0.58	6.6	9.1	-2.5* (1.3)	-28%	0.05
Year 2	4.6	3.8	0.79 (0.52)	21%	0.13	7.7	6.4	1.3 (1.3)	20%	0.31
Cumulative	7.4	7.0	0.42 (0.65)	6.0%	0.52	12	13	-0.60 (1.6)	-4.7%	0.70
Percentage of beneficiaries with any hospital admission with an asthma diagnosis[b]										
Year 1	1.6	1.9	-0.26 (0.34)	-14%	0.44	2.6	3.8	-1.2 (0.87)	-32%	0.16
Year 2	2.1	1.5	0.52 (0.36)	34%	0.14	2.6	2.5	0.08 (0.82)	3.1%	0.92
Cumulative	3.4	3.1	0.31 (0.46)	9.8%	0.51	4.6	5.6	-0.99 (1.1)	-18%	0.36
Percentage of beneficiaries with any ED visits in a time period										
Year 1	33	35	-1.6 (1.2)	-4.7%	0.17	44	49	-4.7** (2.3)	-9.7%	0.04
Year 2	32	33	-0.49 (1.2)	-1.5%	0.68	41	44	-2.7 (2.4)	-6.1%	0.27
Cumulative	49	50	-0.91 (1.2)	-1.8%	0.46	60	63	-2.8 (2.3)	-4.5%	0.22

Table C.1 (*continued*)

	All randomized beneficiaries					Medium- and high-risk beneficiaries only				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Percentage of beneficiaries with any ED visits with an asthma diagnosis[b]										
Year 1	10.1	9.6	0.42 (0.75)	4.4%	0.57	15	15	0.11 (1.7)	< 1%	0.95
Year 2	10.0	10.0	-0.09 (0.79)	< 1%	0.91	16	16	-0.11 (1.8)	< 1%	0.95
Cumulative	16.7	16.1	0.60 (0.94)	3.7%	0.52	25	24	0.79 (2.0)	3.4%	0.70
Sample sizes										
Number of beneficiaries										
Year 1	3,131	3,128				821	819			
Year 2	2,832	2,854				761	759			
Cumulative	3,131	3,128				821	819			

Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2016, to March 31, 2018.

Note: Impact estimates are based on a cross-sectional model of differences in outcomes between treatment and control after randomization, adjusting for demographic characteristics, risk tier, targeted conditions, and CDPS score at randomization as well as baseline expenditures and utilization.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] Asthma-related hospitalizations and ED visits were identified based on all available diagnosis codes on the inpatient and ED claims.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; CDPS = chronic illness and disability payment system; FFS = fee-for-service; PBPM = per beneficiary per month; UIC = University of Illinois at Chicago.

Table C.2. Estimated impact of the UIC intervention on Medicaid expenditures and utilization measures after topcoding at the 98th percentile

	All randomized beneficiaries					Medium- and high-risk beneficiaries only				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)										
Year 1	115	117	-1.7 (4.3)	-1.4%	0.70	154	165	-10 (10)	-6.3%	0.31
Year 2	108	110	-1.3 (4.4)	-1.2%	0.76	136	147	-11 (10)	-7.7%	0.28
Cumulative	117	118	-1.5 (4.2)	-1.3%	0.72	154	165	-12 (9.8)	-7.0%	0.24
Acute inpatient expenditures ($ PBPM)										
Year 1	5.2	5.7	-0.47 (0.65)	-8.3%	0.47	9.4	12	-2.9* (1.7)	-24%	0.08
Year 2	5.7	5.2	0.47 (0.68)	9.0%	0.49	9.2	9.5	-0.37 (1.6)	-3.9%	0.82
Cumulative	8.3	8.2	0.12 (0.78)	1.5%	0.88	14	16	-2.1 (1.9)	-13%	0.27
Outpatient ED expenditures ($ PBPM)										
Year 1	7.8	8.1	-0.26 (0.42)	-3.2%	0.53	12	14	-1.4 (1.0)	-10.0%	0.17
Year 2	8.3	7.9	0.39 (0.47)	5.0%	0.40	13	12	0.78 (1.1)	6.6%	0.47
Cumulative	8.5	8.4	0.12 (0.39)	1.5%	0.75	13	13	-0.04 (0.93)	< 1%	0.96
Pharmacy expenditures ($ PBPM)										
Year 1	23	24	-1.0 (1.2)	-4.3%	0.40	27	31	-3.8 (2.7)	-12%	0.15
Year 2	21	22	-1.6 (1.3)	-7.3%	0.19	25	30	-4.9* (2.8)	-16%	0.09
Cumulative	22	24	-1.4 (1.2)	-5.9%	0.24	27	31	-4.6* (2.7)	-15%	0.09
All other Medicaid expenditures ($ PBPM)										
Year 1	62	63	-0.84 (1.9)	-1.3%	0.66	77	77	-0.16 (4.3)	< 1%	0.97
Year 2	57	58	-1.00 (2.0)	-1.7%	0.61	65	68	-3.3 (4.4)	-4.8%	0.46
Cumulative	61	62	-0.46 (1.8)	< 1%	0.80	75	75	-0.66 (4.0)	< 1%	0.87
Hospital stays, per 1,000 beneficiaries										
Year 1	46	48	-1.9 (6.0)	-4.0%	0.75	83	106	-23 (16)	-22%	0.15
Year 2	53	43	10 (6.3)	23%	0.11	88	79	8.7 (16)	11%	0.58
Cumulative	52	48	3.7 (5.0)	7.7%	0.46	88	101	-13 (14)	-12%	0.36

Table C.2 (*continued*)

	All randomized beneficiaries					Medium- and high-risk beneficiaries only				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED visits, per 1,000 beneficiaries										
Year 1	537	560	-23 (22)	-4.1%	0.30	849	945	-96* (54)	-10%	0.07
Year 2	547	540	7.6 (24)	1.4%	0.75	827	825	1.9 (57)	< 1%	0.97
Cumulative	558	566	-8.2 (19)	-1.4%	0.67	873	927	-55 (48)	-5.9%	0.25
ED visits with an asthma diagnosis, per 1,000 beneficiaries[b]										
Year 1	127	122	4.8 (10)	3.9%	0.63	197	198	-1.4 (24)	< 1%	0.95
Year 2	127	123	3.8 (10)	3.1%	0.72	196	200	-3.1 (24)	-1.6%	0.90
Cumulative	134	129	5.3 (8.9)	4.1%	0.55	210	213	-2.6 (21)	-1.2%	0.90
Sample sizes										
Number of beneficiaries										
Year 1	3,131	3,128				821	819			
Year 2	2,832	2,854				761	759			
Cumulative	3,131	3,128				821	819			

Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2016, to March 31, 2018.

Note: Impact estimates are based on a cross-sectional model of differences in outcomes between treatment and control after randomization, adjusting for demographic characteristics, risk tier, targeted conditions, and CDPS score at randomization as well as baseline expenditures and utilization. 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the 4 semi-annual periods covering the 2 follow up years.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] Asthma-related hospitalizations and ED visits were identified based on all available diagnosis codes on the inpatient and ED claims.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; CDPS = chronic illness and disability payment system; FFS = fee-for-service; PBPM = per beneficiary per month; UIC = University of Illinois at Chicago.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for the Board of Trustees at the University of Illinois, Chicago (UIC) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to UIC. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on three core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for UIC led to a Bayesian estimate of the program's impact on total Medicaid expenditures of between -1 and -2 percent (an estimated reduction of between $2 and $3 per beneficiary per month) in the first two years for the full sample; among higher-risk beneficiaries the impacts were greater at between -10 and -11 percent, or reductions of $21 to $25 per beneficiary per month in the first two years of the program.

Table D.1. Comparison of frequentist and Bayesian impact estimates for UIC in the first two years after enrollment

			Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Sample	Follow-up period	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	Full sample	Year 1	-5.8 (-39, 28)	-3.2 (-16, 9.5)	-3%	-4%	-2%
		Year 2	-6.3 (-44, 31)	-1.9 (-14, 10)	-2%	-4%	-1%
	Higher-risk	Year 1	-42 (-92, 8.5)	-25 (-46, -3.5)	-10%	-19%	-11%
		Year 2	-47 (-99, 4.7)	-21 (-40, -1.6)	-9%	-23%	-10%
Hospital admissions	Full sample	Year 1	-0.16 (-17, 16)	-1.00 (-5.6, 3.5)	-3%	> -1%	-2%
		Year 2	8.5 (-6.6, 24)	-0.51 (-4.6, 3.4)	-2%	18%	-1%
	Higher-risk	Year 1	-28 (-77, 20)	-14 (-25, -1.5)	-9%	-22%	-11%
		Year 2	-5.3 (-47, 37)	-9.6 (-19, -0.47)	-9%	-5%	-10%
ED visits	Full sample	Year 1	-22 (-72, 29)	-14 (-65, 37)	-3%	-4%	-2%
		Year 2	16 (-39, 71)	-8.7 (-58, 39)	-3%	3%	-2%
	Higher-risk	Year 1	-106 (-240, 29)	-118 (-218, -20)	-10%	-10%	-11%
		Year 2	-3.6 (-139, 132)	-94 (-181, -9.4)	-9%	> -1%	-10%

Table D.1 (*continued*)

Source: Mathematica's analysis of information from from the awardee's randomization file and Illinois Medicaid claims and encounter data as of August 10, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

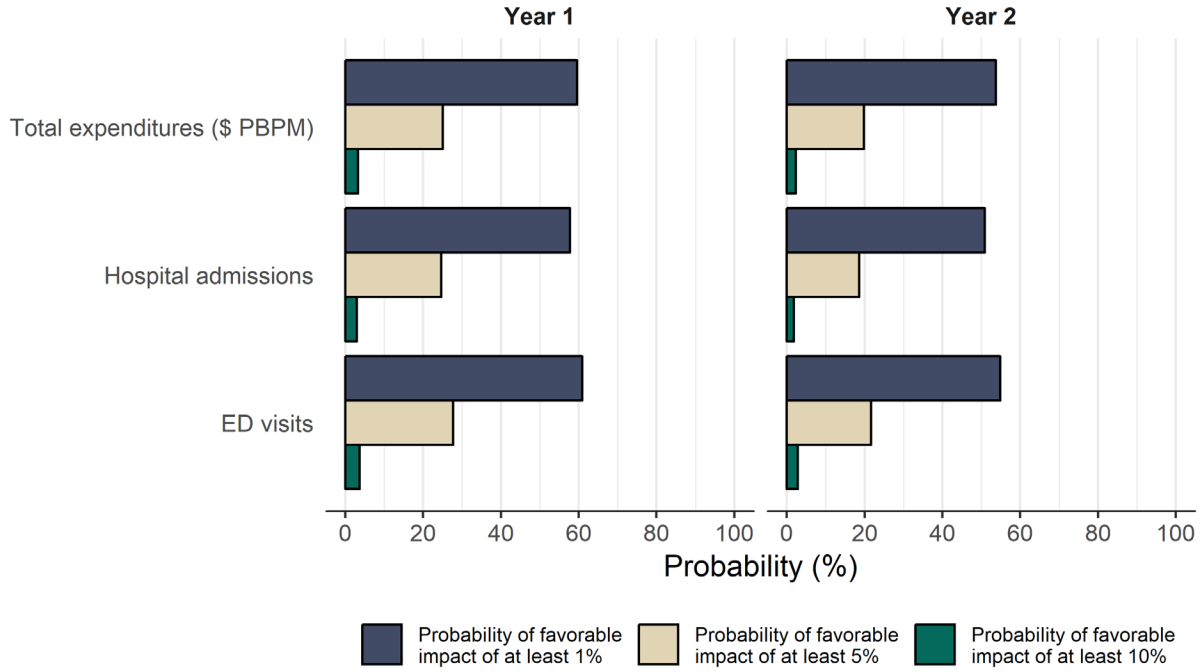
[a] The frequentist and Bayesian results are based on data truncated at the 98th percentile. In the revised draft, this will be updated to use non-truncated data, consistent with results reported in the main body of this report.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Conceptually, the Bayesian model averages the frequentist results and the prior, weighting each component proportional to its precision; for the higher-risk subgroups, where the frequentist estimates are less precise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that most impacts are statistically indistinguishable from zero for the full sample, while the impacts for higher-risk beneficiaries are more promising, if not always statistically distinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figures D.1 and D.2 show, separately for the full and higher-risk samples, the probability that UIC achieved favorable impacts during each of the first two years on three core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the UIC program had a favorable impact on key outcomes for all randomized beneficiaries



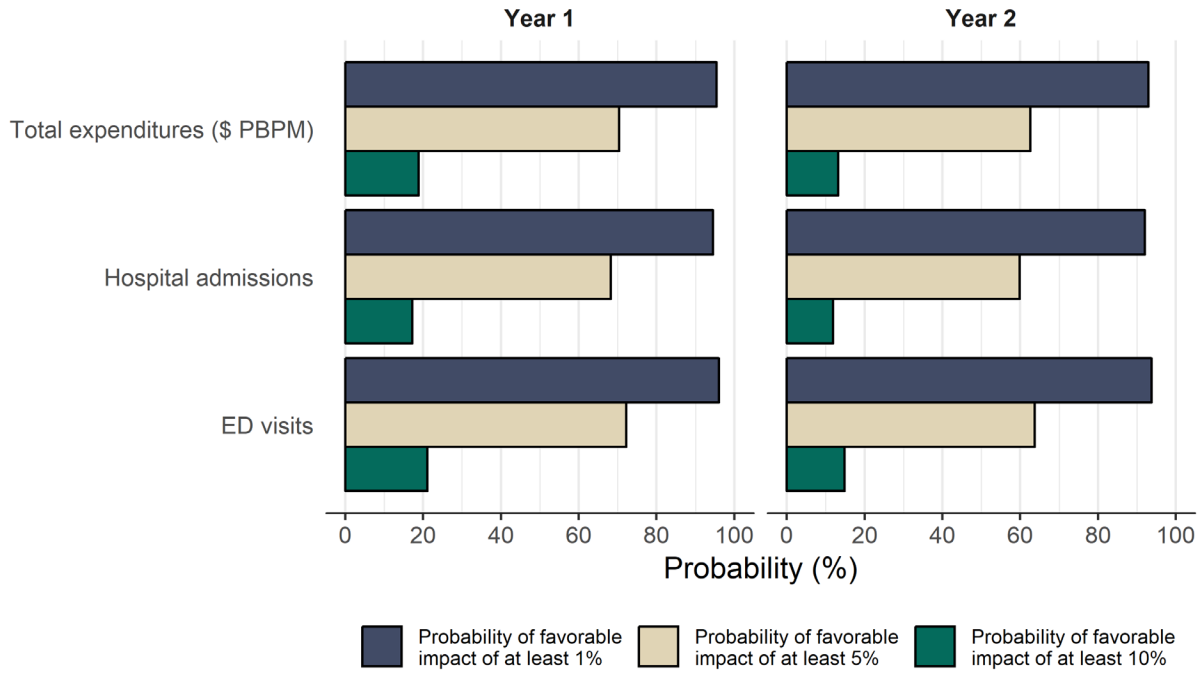
Source: Mathematica's analysis of information from from the awardee's randomization file and Illinois Medicaid claims and encounter data as of August 10, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a modest probability—in the range of 50-60 percent—that UIC had a favorable impact of 1 percent or more on total Medicaid expenditures, hospital admissions, and emergency department visits. The probabilities of an impact of 5 percent or more are smaller at around 25 percent in Year 1 and 20 percent in Year 2; these probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the UIC program did not have a meaningful impact on total expenditures or service utilization in the full sample, despite some promising results.

Figure D.2. Probability that the UIC program had a favorable impact on key outcomes for higher-risk beneficiaries



Source: Mathematica's analysis of information from the awardee's randomization file and Illinois Medicaid claims and encounter data as of August 10, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Impact probabilities are higher in the higher-risk sample, with strong (90 percent) probabilities of an impact of 1 percent or more in both program years and modest (60-70 percent) probabilities of an impact of 5 percent or more for all three outcomes in both program years. These probabilities suggest more substantial impacts concentrated among higher-risk beneficiaries, in line with the more promising results for these subpopulations from the frequentist analysis.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Boston Medical Center

September 2020

Anna Christensen, Dana Petersen, Danielle Chelminsky, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

BOSTON MEDICAL CENTER

Boston Medical Center (BMC), together with its implementing partner, Baystate Medical Center, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Collaborative Consultative Care Coordination (4C) program. The program sought to improve care coordination for children and adolescents who had complex medical conditions. The target population consisted of children diagnosed with at least one chronic condition in any of nine categories (neuromuscular, respiratory, cardiovascular, renal, hematologic, immunologic, metabolic, autism spectrum, and congenital defect), and had high service use in the year before enrollment or were considered to be at risk for high service use. Children were eligible regardless of insurance status or type. The awardee launched the 4C program in December 2014. The intervention period funded by HCIA R2 ended in August 2017. BMC received a four-month no-cost extension through December 31, 2017. The awardee ended enrollment in August 2017 and used the extension to complete data analysis and continue sustainability efforts. Table 1 summarizes the key characteristics of the program.

Important issues for understanding the evaluation

- The 4C program enrolled children diagnosed with at least one chronic medical condition and who had high service use in the year before enrollment, or who were at risk for high service use. The program enrolled 365 children, 252 of whom were Medicaid beneficiaries.
- Due to the small study sample, inability to replicate the eligibility criteria in claims, and the lack of identifiers for some participants, it was not possible to conduct a rigorous impact evaluation of this program.
- Because the program could not supply identifiers for some participants, it was not possible to measure program outcomes in claims data. The quantitative findings in this report are therefore limited to the demographic characteristics of the Medicaid participants as provided by the awardee.

The awardee hypothesized that having multidisciplinary teams provide comprehensive and personalized care coordination to children with medical complexity (CMC), under the direction of complex care pediatricians, would result in better health and health-related outcomes. The 4C program included adopting new software and consumer-facing technology to support its care coordination activities. It also offered consultation services to primary care providers (PCPs), who are often not trained to or do not have the resources to support CMC. The goals of the 4C program were to (1) improve care planning and coordination for CMC, (2) reduce stress and depression among the caregivers of CMC, and (3) lower the cost of care by reducing hospitalizations among CMC.

A rigorous impact evaluation of this program was not possible for three reasons. First, the 4C eligibility criteria could not be replicated in health care claims data because clinical judgment was required to identify one the of targeted groups: children at risk for high service use. Therefore, it was not possible to identify from claims data a comparison group that met the 4C eligibility criteria. Second, the awardee did not collect participants' Social Security numbers,

which limited the ability to link participants to Medicaid data. The inability to link many participants to claims data restricted the ability to measure changes in participants' health care service use and spending over time. Third, because of the lack of patient identifiers, only 152 (60 percent) of the 252 Medicaid participants could be linked to Medicaid administrative data. As a result, the available sample for the study was too small to be able to identify statistically significant impacts of reasonable magnitude for the 4C program. As a result, the descriptive results in this report cannot be interpreted as representing casual impacts of the program.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	BMC implemented the 4C program to provide technology and consultation support to the PCPs of CMC and to help the CMC and their families coordinate social, behavioral, and medical services.
Major innovation	The 4C program was innovative because it offered a greater intensity of care coordination supports and access to care teams with a broader array of disciplines than was previously available, and it emphasized consultation with the PCPs of the CMC. The use of an Internet-based, shared care plan co-developed by the care team and the family was also a key feature of the program.
Program components	• Care coordination • Health information technology
Target population	The 4C program defined CMC as children and adolescents diagnosed with at least one chronic medical condition in any of nine categories (neuromuscular, respiratory, cardiovascular, renal, hematologic, immunologic, metabolic, autism spectrum, and congenital defect) and who had high service use in the year before enrollment or who were considered by the 4C staff to be at risk for high service use. CMC could enroll in the 4C program regardless of insurance status or type.
Total enrollment	BMC enrolled a total of 365 children (81 percent of its original enrollment goal). Among them, 252 participants (69 percent) were Medicaid beneficiaries. Participants in Medicaid ranged in age from birth to 21 years, with an average age of 8 years. Nearly 60 percent were male. Forty-two percent (107) were enrolled at BMC and 58 percent (145) were enrolled at Bayside Medical Center.
Theory of change or theory of action	The awardee hypothesized that improving care coordination for CMC would lead to increased access to appropriate services and supports for CMC and their families, which would decrease caregivers' stress and depression and improve the CMC's health. Better CMC health would lead to fewer hospital stays and thus decreased health care costs.
Award amount	$6,128,059
Effective launch date	The program began in December 2014, three months after award date.
Program settings	Multidisciplinary care teams, including a nurse care coordinator, met with participants and their families in the clinic during a comprehensive intake assessment. There were follow-up appointments in the clinic one month after intake and every six months thereafter, at minimum. The coordinators provided other support over the phone and in participants' homes or in other community settings, including schools and community-based health centers. They also provided care coordination through health technology, via the cloud-based Internet portal available to the participants, their families, and their health care providers.
Market area	Boston and Springfield, Massachusetts

Table 1 (*continued*)

Program characteristics	Description
Target outcomes	• Reduce caregivers' stress and depression • Fewer hospital admissions • Lower total cost of care
Payment model	At the end of the award, both BMC and Baystate Medical Center were negotiating payment approaches with local Medicaid ACOs to fund and continue the 4C program. As of August 2018, eight months after the extended award end date, neither site had reached an agreement with an ACO.
Sustainability plans	At the end of the award, BMC and Baystate Medical Center ended services to children in the 4C program. Program staff transferred care coordination responsibilities to primary care and auxiliary providers involved in the children's care.

ACO = accountable care organization; BMC = Boston Medical Center; CMC = children with medical complexity; 4C = Collaborative Consultative Care Coordination program; PCP = primary care provider.

PROGRAM DESIGN AND ADAPTATION

The 4C program had two key components: care coordination and health technology.[1]

Care coordination

Multidisciplinary care teams conducted an initial two-hour, in-person intake assessment of each child's medical history and current health status; health care and health care service needs (including physical, developmental, and behavioral health); and the family's social needs. The care teams included a complex care pediatrician, nurse care coordinator, social worker, family navigator, dietician, and a psychiatrist, psychologist, and/or developmental behavioral physician. Based on this thorough assessment, the nurse care coordinator developed a care plan that outlined the child's and family's medical and social goals and needs. The care teams then worked with the child and family to coordinate care and supports across specialty medical providers and social services to achieve the goals and remedy the medical and social issues identified in the care plan. This included, at a minimum, conducting follow-up appointments with the participant and family one month after the intake assessment and every six months thereafter until the end of the award. The care teams also interacted with participants via telephone calls and as-needed in-person visits to provide services such as (1) making referrals to and assisting families with scheduling appointments with other providers, (2) attending appointments together with families, (3) acquiring medical supplies for the children and teaching families how to use those supplies, (4) assisting with access to special services at schools, (5) finding housing or food supports, (6) helping families overcome transportation challenges, and (7) helping families to be more self-sufficient in addressing their needs. The frequency of contact with the program varied based on participants' needs.

[1] The Third Annual Evaluation report provides additional details on the design and implementation of the 4C program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Health technology

Program staff used a secure, cloud-based, Internet portal called ACT.md to make the care plans available to families, PCPs, and other providers involved in the children's care, as well as to share information and track care coordination activities. At the one-month follow-up appointment, the nurse care coordinator reviewed the care plan with the family, obtained approval of the plan, published the care plan in the portal, and trained the family to use the portal to access the care plan and communicate with their care team.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee faced challenges meeting its enrollment goal. They initially planned to identify and recruit CMC with high service use by accessing and analyzing medical claims data from insurers. The awardee was unable to overcome payers' privacy concerns regarding the sharing of necessary claims data. As a result, in April 2015, the awardee expanded the 4C eligibility criteria to allow enrolling children at risk for high service use based on clinical judgement, rather than enrolling only those with documentation of high service use (which the program had defined as 10 or more clinic visits or 10 or more hospital days in a year). The new eligibility group of children at risk for high service use included, among others, children with any complicating psychosocial and economic factors that were (or were at risk of) adversely affecting outcomes, including children whose caregivers had significant stressors.

The awardee reported that it delivered services as intended and on schedule for most participants. However, staff felt challenged in later years with providing the same level of care as in the first program year because their caseloads grew as the pace of enrollment increased. In a survey of 4C program staff conducted at the start of the third program year, 60 percent reported that their participant caseload was too heavy and 40 percent reported that they had insufficient time for the amount of work they wanted to do.

Implications of program implementation for achieving program goals

- The small number of Medicaid participants and the use of clinical judgment to identify children eligible for the program made it impossible to evaluate the impact of the 4C program on service use and expenditures.
- As their caseloads grew, program staff had difficulty maintaining the intensity of care coordination services originally planned.

Although the awardee was successful in its initial engagement of participants' families, maintaining their long-term commitment to the program and encouraging families' use of the ACT.md Internet portal to access the care plan was more challenging. Despite having received tutorials on how to use the system, many families lacked the technological literacy or English-language proficiency to use the platform effectively or were too overwhelmed with other competing demands to prioritize using the system. Almost half of the 4C staff surveyed reported

that the program required too much time of participants and had too many requirements for the families.

The awardee also had difficulty engaging participants' PCPs and other providers in care coordination activities and in using the ACT.md care coordination software. The awardee intended for providers outside the program to communicate with 4C staff about the participants' needs and access the children's care plans on ACT.md. Staff said that some external providers were not interested in partnering with 4C staff to coordinate care and some were simply difficult to reach, likely due to competing priorities. Still other providers coordinated with 4C staff outside of ACT.md (for example, by phone) but did not want to learn to use another software system for communication.

Despite these challenges, program leaders and staff felt that 4C had a positive effect on the delivery of care for CMC and likely decreased parental stress, reduced service duplication, and improved children's quality of life. They also felt that the 4C program offered valuable services to CMC and their families. In a survey of 4C staff, 100 percent of staff reported that the program had a positive impact on participants' satisfaction and quality of life. However, the evaluation was unable corroborate staff perceptions by empirically measuring the impact of the program on these outcomes.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Enrolling and engaging participants

The awardee aimed to enroll 450 participants by the end of the cooperative agreement. To identify CMC with or at risk of high service use, 4C staff invested in multiple outreach and recruitment efforts, including communicating directly with PCPs in the community to identify patients who qualified; meeting with families of potential participants during their hospital, neonatal intensive care unit, or pediatric intensive care unit stays; advertising at community health centers, schools, and social service agencies; and conducting other community outreach, such as presentations. The 4C program enrolled participants after they were confirmed as eligible through medical record review or clinical judgment, completed the two-hour in-person intake assessment, and signed a consent form.

After a slow start to enrollment during the first program year and steady progress in the second program year, the awardee came close to meeting its enrollment goal in the final year of the cooperative agreement, enrolling 365 participants (81 percent of its original enrollment goal) from December 2014 through August 2017. Children could enroll in the 4C program regardless of insurance status or type.

According to 4C leaders and staff, the program was mostly successful in engaging and delivering services to participants as intended. Staff reported performing intake assessments and developing

care plans for all enrolled CMC. They reported conducting at least the minimum required follow-up visits at one month after intake and every six months thereafter for most participants. The awardee reported that, as of May 2017 (three months before the program ended), 4C staff had provided more than 15,000 unique encounters to the 365 participants since the program's launch, for an average of 41 encounters per child. For example, during the third quarter of the last program year (March through May 2017), these encounters included 852 telephone contacts, 266 in-person appointments, and 107 online or email contacts.

Characteristics of program participants

Of the 365 children who enrolled in the 4C program from December 2014 through August 2017, 252 (69 percent) were covered by Medicaid or the Children's Health Insurance Program (CHIP). (The demographic information provided by the awardee included only enrollees covered by Medicaid or CHIP.) Medicaid participants ranged in age from birth to 21 years and were an average age of 8 years at the time of enrollment. Nearly 60 percent of all Medicaid participants were male. More than half (55 percent) of the Medicaid participants enrolled through Baystate Medical Center. The rest of the Medicaid participants enrolled through BMC.

Challenges of measuring program impacts

Due to the small number of Medicaid enrollees and an inability to replicate eligibility criteria (one of which required clinical judgment) in claims data, it was not possible to conduct a rigorous impact evaluation of this program. In addition, the awardee did not collect many patient identifiers (such as Social Security numbers), which meant that few 4C participants could be linked to and identified in Medicaid enrollment and claims data. The lack of participant identifiers also limited the ability to describe health care use of the 4C participants. As a result, the findings in this report were limited to a descriptive analysis of the demographic characteristics of program participants based on the awardee's data.

CONCLUSION

BMC and its partner Baystate Medical Center were partially successful in implementing the 4C program to provide intensive care planning and care coordination to CMC in Massachusetts. The program provided personalized, intensive services that went beyond the care coordination previously offered at those sites. Implementation challenges included that the program was time intensive for families to remain engaged and use ACT.md, it was difficult for care coordinators to provide the level of highly personalized services they wanted to as their caseloads grew, and the care coordinators had difficultly engaging the participants' providers who were outside of the 4C program. Despite these challenges, program staff perceived that the 4C program had positive impacts on care delivery, caregivers' stress level, and health and social outcomes for the children and their families. However, the small number of participants and the use of clinical judgment to determine program eligibility meant that the program's impacts on outcomes could not be rigorously assessed. The program's decision not to collect patients' Social Security numbers and

inability to provide national Medicaid identification numbers for many patients also precluded comparison of pre- and post-program outcomes for the enrollees.

PROGRAM SUSTAINABILITY

BMC and Baystate Medical Center ended services to 4C program participants when their award ended in December 2017. The awardee stopped enrolling new participants in August 2017, but continued serving its existing enrollees during the four-month extension period and sought to ensure that participants' care coordination needs would be met after the program ended. Program staff transferred care coordination responsibilities to the children's PCPs or other providers involved in the children's care, for example early intervention and special education staff at schools, and staff affiliated with providing durable medical equipment and training.

BMC proposed that it, along with its partner, Baystate Medical Center, negotiate payment approaches with the Medicaid ACOs to fund and continue the 4C program after the end of the award. The awardee focused on ACOs as a source of funding for the 4C program because of a statewide effort in Massachusetts to implement regional Medicaid ACOs starting in January 2018 as a part of the state's five-year innovative Section 1115 Medicaid waiver. The awardee hoped that both sites would reach agreements by March 2018, when the new ACOs were expected to launch. However, neither site had reached agreements with ACOs by August 2018.

BMC's proposed payment model

BMC and Baystate Medical Center negotiated payment approaches with their Medicaid ACOs to continue funding of the 4C program. Each site had flexibility to negotiate its own payment approach with the ACOs. Baystate Medical Center was negotiating a per beneficiary per month fee of $100, which would include all program services except for physician visits. At the award's end, BMC was still determining the staffing model needed to serve the broader ACO population before determining its desired payment approach. By August 2018, neither site had reached an agreement with its ACO.

The awardee reported that both sites were continuing aspects of the 4C programs or were applying lessons learned from the program in new contexts. Some 4C staff secured new care coordination positions in their sites, where they provide services similar to those of 4C to CMC or other populations. The awardee also won a contract to provide wrap-around services for CMC in foster care, for which they said they will use lessons learned about service and staffing models from the 4C program.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: CareChoice Cooperative

September 2020

Poonam Pardasaney, Bob Schmitz, Julia Doherty, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

CARECHOICE COOPERATIVE

CareChoice Cooperative, a cooperative of skilled nursing facilities (SNFs), senior independent housing, and assisted living communities in Minnesota, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the Person-Centered Care Connections (PCCC) program. The program represented an expansion of an earlier CareChoice pilot program, called the Resident-Centered Care Connections (RCCC). Its goal was to improve the care and safety of SNF patients who transitioned to the community and to reduce their total cost of care. The program launched in 10 SNFs in Minnesota in January 2015 and ended in August 2017. Table 1 summarizes the PCCC program's key characteristics.

The awardee hypothesized that the PCCC comprehensive transition planning program would improve patients' and families' understanding of the discharge plan and self-management strategies and improve the coordination of post-discharge follow-up care. In turn, the awardee expected these factors to result in faster recovery, better health, and earlier intervention for health-related problems. The awardee expected that the program would reduce hospital readmissions within 30 days of discharge home by 20 percent and total Medicare spending by 3.5 percent. The awardee also aimed to reduce hospital readmissions within 90 days of SNF discharge to home but did not set a goal for this outcome.

Important issues for understanding the evaluation

- The PCCC program sought to improve the care and safety of SNF patients who transitioned to the community. It represented an expansion of an earlier CareChoice pilot program, the RCCC.
- The program aimed to reduce hospital readmissions during the 90 days following discharge from a SNF to home and thereby reduce total costs of care.
- The intervention group for the evaluation included 900 Medicare fee-for-service (FFS) beneficiaries (among the total enrollment of 8,016) admitted to a participating SNF from January 2015 through March 2017 and discharged home. The comparison group included 2,563 matched Medicare FFS beneficiaries who were admitted to nearby nonparticipating SNFs during the same period and discharged home.

Table 1. Program characteristics at a glance

Program characteristic	Description
Purpose	The purpose of the PCCC was to improve the care and safety of SNF patients who transitioned to the community and reduce readmissions and costs of care for these patients.
Major innovation	The PCCC designed a robust discharge planning process supported by an innovative web-based application called Engage. Engage incorporated components of the hospital-based Re-Engineered Discharge program adapted to the SNF setting. The program combined the Engage software with the introduction of dedicated transition coordinators, interdisciplinary team training, and process improvement strategies to provide safer, better quality of care to SNF patients being discharged to the community.

Table 1 (*continued*)

Program characteristic	Description
Program components	• **Transitional care coordination** included improved communication between interdisciplinary transition team members; comprehensive discharge planning and transition documentation for patients, caregivers, and providers; and post-discharge follow-up phone calls. • **Patient and family engagement** was incorporated during the SNF stay to improve education of participants and families about the medical condition and transition plan. • **Quality improvement and workflow process redesign** was implemented using the web-based decision support tool, Engage, and enhanced transition planning services. • **Staff education and training** was provided to the interdisciplinary team focusing on its role in process improvements to improve discharge planning and care coordination.
Target population	The program sought to engage all SNF patients discharged to the community, regardless of condition or payer. No patient consent was required.
Participating providers	10 CareChoice Cooperative SNFs in five counties in Minnesota participated.
Total enrollment	The awardee enrolled 8,016 patients, representing 94 percent of its original enrollment goal.
Level of engagement	Among all enrollees, the awardee reported that 79 percent had a successful follow-up phone call within 90 days after transition, with 83 to 85 percent having earlier follow-up calls. Almost all patients (92 percent) reported keeping their scheduled appointment with a clinician by the 30-day follow-up phone call.
Theory of change theory of action	The awardee hypothesized that introducing transition coordinators and software to facilitate improved transition planning would result in developing a comprehensive post-discharge continuing care plan shared with patients, families, and home care and primary care providers in the community. Staff training, increased coordination, and enhanced education of patients and families would, in turn, better engage patients and better prepare them to safely transition to the community. Finally, the awardee expected these services to improve participants' and families' satisfaction and decrease hospital readmissions and overall health care costs.
Award amount	$3,347,584
Effective launch date	January 1, 2015
Program setting	SNFs
Market area	Urban, suburban
Market location	Minnesota
Target outcomes	• Increased patient and family understanding of the discharge plan • Increased satisfaction of patients and families • Reduced hospital readmissions • Reduced total cost of care
Payment model	A new FFS payment amount
Sustainability plans	The awardee decided to forgo the originally proposed payment model that relied on FFS payments after its analyses showed the program did not achieve its intended goals of reducing costs and readmissions. The awardee anticipated that, without these outcomes, it would be difficult to negotiate a FFS payment model with payers.

FFS = fee-for-service; PCCC = Person-Centered Care Connections; SNF = skilled nursing facility.

The impact analysis presented in this report was limited to 900 eligible Medicare FFS beneficiaries who were admitted to a participating SNF unit from January 2015 through March 2017, and were discharged home. The comparison group included 2,563 Medicare FFS beneficiaries with similar characteristics who were admitted to nonparticipating SNFs located in the intervention and neighboring counties in Minnesota during the same period and discharged home. Table 2 summarizes the key features of the impact evaluation. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The evaluation estimated program impacts by comparing outcomes for the treatment group to those of the matched comparison group, controlling for pre-enrollment characteristics of the sample members. Assuming that external factors affected the treatment and comparison groups similarly, a comparison group well matched on observable characteristics should produce unbiased estimates of program effects.
Intervention group for evaluation	The intervention group for the evaluation relied on 900 Medicare FFS beneficiaries (among the total enrollment of 8,016) admitted to a participating SNF from January 2015 through March 19, 2017, and discharged home, representing 11 percent of total enrollment during that period. The study intervention group excluded 3,716 enrollees (46 percent) whose payer was Medicaid or a commercial insurer and 2,109 enrollees (26 percent) who were not discharged to the community (that is, they were transferred to another SNF, died, or remained in the SNF). It also excluded 1,127 enrollees (14 percent) who were in a Medicare Advantage plan and 164 enrollees (2 percent) who lacked Medicare Part A or Part B, did not have Medicare as the primary payer, or could not be matched to the Long-Term Care Minimum Data Set needed for the analysis.
Comparison group	The comparison group included 2,563 Medicare FFS beneficiaries who were admitted to nonparticipating SNFs located in the intervention and neighboring counties in Minnesota during the same period and discharged home, and who were matched to the treatment group patients using propensity scores.
Limitations	If local area factors that affect SNF patients' outcomes differed for the treatment and comparison practices, or if the treatment and comparison SNFs have had different outcomes historically, the impact estimates could be biased.

FFS = fee-for-services; SNF = skilled nursing facility.

PROGRAM DESIGN AND ADAPTATION

The PCCC service delivery model had four components embedded into SNFs' enhanced transition planning processes: (1) transitional care coordination, (2) patient and family engagement, (3) quality improvement and workflow process redesign, and (4) staff education and training.[1] The awardee used a web-based decision-support tool, called Engage, to support program implementation and goal tracking. The interdisciplinary care team was comprised of staff working in the SNF unit and its related departments, and included unit nurses, therapists, dieticians, social workers, admissions staff, and medical records staff.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Transitional care coordination

The goal of the transitional care coordination component was to improve the coordination of care with home health and outpatient providers following participants' discharge from SNF. The component consisted of seven services: (1) initiation of comprehensive transition planning on admission; (2) participant surveys on admission and before discharge to assess readiness for transition to community; (3) ongoing needs assessment and targeted support throughout the stay to ensure successful transition to community; (4) assistance with post-discharge appointment scheduling with outpatient providers before participants' discharge; (5) a detailed written transition plan for use at home, describing the reason for admission, services and educational modules delivered in the SNF, post-discharge instructions and medication schedule, and contact information for questions; (6) transfer of the written transition plan to patients' home care and primary care providers and, beginning in the third program year, to patients' pharmacists; and (7) follow-up phone calls with participants and families 48 hours, 30 days, and 90 days after discharge to provide additional care coordination support to connect patients with primary care providers.

Patient and family engagement

The goal of the patient and family engagement component was to educate patients and families about the patient's medical condition, medication regimen, and transition care plan and to help patients and families make informed choices about their care. Engaging patients and families began on SNF admission and continued after discharge. It included four activities: (1) completing a patient survey during the SNF stay to assess needs for discharge home; (2) participating in medication reconciliation with a nurse during the SNF stay; (3) participating in educational modules about self-management of health conditions, medication safety and management, and general wellness during the SNF stay; and (4) participating in post-discharge follow-up calls, as described earlier.

Quality improvement and workflow process redesign

The goal of quality improvement and workflow process redesign was to develop a more robust and systematic approach to discharge planning by introducing new processes, quality improvement activities, supporting software, and designated transition coordinators. The program-based quality improvement and workflow process redesign on the principles of Project RED (Re-Engineered Discharge), a nationally recognized framework for discharge planning and improving workflow processes. The redesigned workflow process emphasized (1) educating patients, (2) assessing patients' understanding, (3) evidence-based discharge planning following national practice guidelines, (4) scheduling post-discharge follow-up and testing appointments, (5) organizing post-discharge services in the community, (6) expediting discharge summary transmission to the primary providers, and (7) conducting post-discharge follow-up calls. The designated transition coordinator at each facility monitored these revised workflow processes using the Engage software tool.

Staff education and training

The goal of staff education and training was to ensure that interdisciplinary teams and newly hired transition coordinators understood the redesigned workflow processes, use of Engage software, and collaborative team approach to improve transitional care coordination. Staff at participating SNFs received training on using the Engage tool Project RED principles, and root cause analyses of acute care transfers and operational challenges. The program combined initial staff training with ongoing support and monthly meetings. During the monthly meetings, PCCC staff reviewed progress and challenges in completing Engage tool components, discussed transition planning barriers and facilitators, and identified additional education and training needs. Further, staff reviewed facilities' self-monitoring reports and action plans, PCCC and Engage monitoring reports, root cause analyses results, and program staff site visit findings. PCCC staff also trained and supported new team members when staffing changes occurred.

Engage tool

The award made the Engage tool available at participating SNFs to provide a more systematic way for interdisciplinary teams to develop and refine comprehensive transition plans for each patient. Align, an organization focused on improving care transitions in post-acute care settings, created the tool. Align designed Engage as a patient transition tool that provided a transparent view of patients' progress as they moved from admission through care in the SNF and transitioned home. The Engage tool assisted interdisciplinary teams with tracking and completing a robust transition planning process, beginning with SNF admission and including learning lessons with patients and their families. The interdisciplinary teams used the tool to review their progress in systematically developing each participant's transition plan and recording completion of required tasks. These entries related to patients' education, medication, and other services provided became part of the final discharge plan shared with patients, families, and post-discharge home health and outpatient providers. The software helped the teams and program staff track and report metrics and identify areas for improvement throughout participants' SNF stays.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Participating SNFs hired the necessary staff to implement the PCCC program and began staff training before program launch. The program met operational milestones, demonstrating early success partly due to participating SNFs' involvement in the prior RCCC pilot that aimed to help facilities reduce avoidable hospital readmissions. CareChoice reported successfully delivering intervention services as intended and meeting most patient satisfaction measures throughout the cooperative agreement. Most SNFs achieved their internal monthly targets soon after program launch on such measures as the percentage of patients who received follow-up calls following discharge, went home with transition plan documents, had transition plans sent to the primary physician, and kept appointments with their community physicians. In addition to meeting

several self-monitoring goals within the first program year, participating facilities also showed consistent and continued improvement throughout the three-year cooperative agreement. Participants reported high levels of satisfaction with the services received. About 75 percent of participants, on average, reported nursing home staff explained things in a way they easily understood and that the transition plan was useful when the participant got home. In addition, 98 percent on average reported they would recommend the facility to others.

Implications of program implementation for detecting impacts

- The significant number of concurrent national, state, and regional initiatives related to nursing home quality improvement and readmissions reduction might have limited the ability of the PCCC intervention to improve outcomes more than other initiatives.
- All 10 facilities participated in CareChoice's three-year precursor initiative designed to reduce avoidable hospital readmissions and enable effective care transitions. The earlier initiative ended in 2013 and could have made it difficult for the PCCC intervention to further improve outcomes.
- The awardee reported a significant unforeseen increase in overall patient acuity during the program compared with the baseline period, which might have affected the ability to achieve projected reductions in readmissions and costs.

Program leaders reported some challenges meeting staff training needs, with staff noting they would have benefited from more intensive, earlier hands-on training on the Engage tool and its advanced tracking capabilities, rather than learning them while on the job. The awardee also faced challenges with staff turnover at participating facilities, which resulted in some SNFs not meeting their program monitoring targets when they hired and trained replacement positions. To address this problem, PCCC program leaders worked with each SNF to create transition coordinator sustainability plans and closely monitored facility performance when staff turnover occurred. CareChoice also encountered lack of Engage software interoperability with some SNFs' electronic medical record systems, requiring duplicative data entry by staff into different platforms for the duration of the cooperative agreement.

ESTIMATING PROGRAM IMPACTS

Recruiting, enrolling, and engaging participants

The program automatically enrolled in the program and delivered program services to all patients admitted to participating SNFs during the intervention period, regardless of condition or payer. Thus, the program required no active recruitment efforts to enroll participants. The awardee reported high levels of patient engagement, with 85 percent of participants having a successful 48-hour post-discharge follow-up call, 83 percent having a successful 30-day call, and 79 percent having a successful 90-day call. Almost all patients (92 percent) had kept their scheduled appointment with a clinician by the 30-day phone call.

Study sample

The awardee reported that 8,016 all-payer patients received PCCC services. The analysis excluded 3,716 enrollees (46 percent) whose payer was Medicaid or a commercial insurer and 2,109 enrollees (26 percent) who were not discharged to the community (that is, they were transferred to another SNF, died, or remained in the SNF). The study also excluded 1,127 enrollees (14 percent) who were in a Medicare Advantage plan and 164 enrollees (2 percent) who lacked Medicare Part A or Part B, did not have Medicare as the primary payer, or could not be matched to the Long-Term Care Minimum Data Set needed for the analysis. These exclusions left an analysis sample of 900 Medicare FFS enrollees (11 percent of all enrollees) with the required data. Appendix A, Table A.1 shows the identification of the final analytic sample from all program participants.

The comparison group included 2,563 matched beneficiaries admitted to 46 nonparticipating SNFs from January 1, 2015, through March 19, 2017, and discharged home. The 46 comparison SNFs were in the intervention counties or selected neighboring counties in Minnesota. To match treatment SNF characteristics, comparison SNFs had to be nonprofit or government-owned, Medicare-certified, and nonhospital-based entities. Nonparticipating CareChoice SNFs were eligible for inclusion in the comparison group if they met these criteria, because the awardee indicated there was no spillover of the intervention to these facilities.

Characteristics of treatment and comparison group beneficiaries

A comparison of treatment and comparison group baseline characteristics confirmed that the two groups were well balanced (Table 3). The average age of treatment and comparison group members at enrollment was 80 years, with the most being 75 years or older. A large majority (92 percent) of the treatment and comparison samples were White. Most treatment and comparison group members (84 percent) became eligible for Medicare on the basis of age and 14 percent were dually eligible for Medicare and Medicaid. The mean hierarchical condition category (HCC) score for both treatment and comparison beneficiaries was more than twice the national average for Medicare beneficiaries. Because the evaluation restricted the analytic sample to beneficiaries with an acute care discharge in the 30 days before SNF admission, the proportion of patients with any hospitalization in the baseline year is 100 percent. Appendix B provides the full balance results measured during the 12 months before enrollment.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Measure	Comparison group (N = 2,563)	Treatment group (N = 900)
Demographics		
Age at enrollment, years	80	80
Age group, %		
Younger than 65	8.2	9.1
65 to 74	18	18
75 to 84	32	34

Table 3 (*continued*)

Measure	Comparison group (N = 2,563)	Treatment group (N = 900)
85 and older	42	40
Female	66	67
White, %	92	92
Original reason for Medicare eligibility, %		
Old age and survivor's insurance	84	84
Disability insurance benefits	16	16
ESRD	0.33	0.34
Medicare/Medicaid dual status, %		
Medicare/Medicaid dual status	14	14
HCC score[a]		
Mean	2.5	2.5
Standard deviation	1.5	1.4
Service use and expenditures during the year before enrollment		
Any hospitalizations, %	100	100
Any ED visits, %	41	41
Number of hospital admissions (per 1,000)	1,457	1,462
Number of outpatient ED visits (per 1,000)	767	790
Total Medicare expenditures ($ PBPM)	2,490	2,530

Source: Mathematica's analysis of SNF admissions from January 1, 2015, through March 19, 2017 and information from Medicare claims and enrollment data as of November 2017.

Notes: The study defined the baseline year as the 365 days before each treatment beneficiary's enrollment date and comparison beneficiary's pseudo-enrollment date. It defined the enrollment or pseudo-enrollment date as the SNF admission date. All beneficiary characteristics were measured during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

None of the differences between treatment and comparison group participants in any of the baseline characteristics differed statistically from zero at the 0.10 level, 2-tailed test.

Appendix B presents the full balance results. An exact match was required for quarter of SNF admission.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact estimates were obtained from a regression comparing differences in post-implementation outcomes between the treatment and comparison groups. This cross-sectional, post-period approach was necessary because the relationship between outcomes and regressors in the pre-intervention period (when all sample members were hospitalized and admitted to a SNF) would be quite different from their relationship in the follow-up period (when all sample members returned to the community). If unobserved factors specific to the individual SNFs affected the treatment and comparison groups similarly, a comparison group well matched on

observable characteristics should produce unbiased estimates of program effects under this cross-sectional, post-period design.

The study defined the enrollment date (or pseudo-enrollment date for comparison group beneficiaries) as the SNF admission date. It defined the pre-enrollment period for assessing baseline characteristics as the year before each participant's enrollment date (or pseudo-enrollment date for comparison group beneficiaries). The study also defined the post-enrollment period as the two years after the enrollment date. The post-enrollment period included costs related to the precipitating SNF stay, because the program initially engaged participants while they were in the SNF.

The core outcomes for the study were total Medicare spending, number of hospital admissions, and number of ED visits, assessed in the first and second follow-up years. The cross-sectional regression model controlled for beneficiaries' demographic characteristics, clinical characteristics at baseline and SNF admission, and SNF-level cost and use, case mix, star rating, and geographic characteristics. Appendix A describes in detail the statistical models and the analytic sample used to estimate the effects of the program.

IMPACT RESULTS

The PCCC program did not have a discernible impact on Medicare spending or hospital readmissions during the first two years after enrollment (Table 4). However, although the theory of action did not target emergency department (ED) visits, the intervention was associated with a statistically significant estimated **21 percent** reduction in ED visits among treatment group beneficiaries relative to the comparison group in the first follow-up year. This effect persisted in the second follow-up year. The lower estimated number of ED visits in the treatment group did not result in lower outpatient payments or higher primary care or specialist visits than in the comparison group (data not shown; see Appendix C). Appendix D shows the results from the Bayesian analysis.

Table 4. Estimated impact of the PCCC intervention on selected outcomes

Outcome	Year 1	Year 2
Total expenditures ($ PBPM)		
Impact ($)	164	109
Percentage impact	6.0%	7.8%
p-value	0.48	0.46
Hospital stays, per 1,000 beneficiaries		
Impact (rate)	13	42
Percentage impact	1.9%	8.4%
p-value	0.86	0.51
Hospital readmissions, per 1,000 discharges		
Impact (rate)	-0.03	-0.89

Table 4 (*continued*)

Outcome	Year 1	Year 2
Percentage impact	< 1%	-5.4%
p-value	0.98	0.74
ED or observation visits, per 1,000 beneficiaries		
Impact (rate)	-212***	-160*
Percentage impact	-21%	-20%
p-value	< 0.01	0.05
Sample sizes		
Number of beneficiaries		
Treatment	900	715
Comparison	2,563	1,979
Number of index hospital discharges for readmissions		
Treatment	1,403	316
Comparison	4,019	918

Sources: Mathematica's analysis of SNF admissions from January 1, 2015, through March 19, 2017, and information from Medicare claims and enrollment data as of November 2017.

Note: Impact estimates for all outcomes represent regression-adjusted treatment–comparison differences based on a post-period cross-sectional regression that controls for beneficiaries' characteristics and the value of the outcome variable at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month; PCCC = Person-Centered Care Connections; SNF = skilled nursing facility.

The lack of estimated program impacts on hospital readmissions and total costs is surprising given the program's successful implementation, the supporting evidence base for the intervention, and the high levels of staff and participant engagement. Further, the treatment and comparison group beneficiaries were well matched at baseline and the analysis controlled for a comprehensive set of beneficiary- and SNF-level characteristics that could affect the outcomes. The awardee reported that the lack of anticipated effects could be related to a significant unforeseen increase in overall SNF patients' acuity during the program period compared with the baseline period, which might have affected the ability to achieve the projected reductions in readmissions and costs. Lack of advanced training on the Engage tool and its software interoperability challenges might have made it more difficult for the awardee to meet its outcome goals as well. In addition, CareChoice piloted the RCCC program, a precursor to the

Main findings from impact evaluation

- The PCCC intervention had no estimated impact on the target outcomes of total Medicare spending and hospital readmissions.
- A persistent estimated reduction in ED use was observed in the follow-up period.

PCCC program, from mid-2010 to mid-2013. All 10 participating SNFs (and likely some comparison facilities) also participated in this pilot program. The existence of this program, which predated the baseline period for this study (2014), could have affected the ability to reduce the incremental effectiveness of the CareChoice intervention. It is possible that PCCC participants were entering a facility whose mean outcomes had already been affected favorably by the RCCC program. However, without the resources required to continue the program, it seems unlikely that any effects of the pilot program on this subset of comparison SNFs would have persisted in the new program period, or had much influence on the overall comparison group mean.

CONCLUSION

There is no evidence that the PCCC intervention affected the targeted outcomes of total Medicare spending and post-discharge hospital readmissions. Although the intervention was associated with fewer ED visits during the follow-up period, there was no corresponding decrease in outpatient payments nor increase in primary care or specialty provider visits to suggest that the program changed patterns of ambulatory care service use. The lack of a treatment effect might relate to concurrent initiatives and technical assistance provided to all Minnesota SNFs that could have reduced the estimated effects of the PCCC program relative to the comparison group. The PCCC program began when national, statewide, and regional initiatives and regulatory agencies focused on improving care transitions, reducing hospital readmissions, and improving nursing home quality and efficiency.[2,3,4,5] Nursing homes faced regulations requiring implementation of Quality Assurance and Performance Improvement Programs and increased incentives to reduce hospitalizations. At the same time, many SNFs participated in Minnesota's three-year performance improvement plan with similar goals as the PCCC program. These coexisting campaigns, recent initiatives, and regulatory factors in the nursing home environment likely benefitted patients of both treatment and comparison SNFs and reduced the ability to detect differential impacts associated with the PCCC intervention.

Limitations of evaluation

The evaluation of the PCCC program faced two limitations, which might have resulted in an underestimation of program effects. First, the involvement of some CareChoice SNFs in the precursor pilot intervention might have limited the ability to detect a potential program effect.

---

[2] Popejoy, L.L., A.A. Vogelsmeier, B.J. Wakefield, C.M. Galambos, A.M. Lewis, D. Huneke, and D.R. Mehr. "Adapting Project RED to Skilled Nursing Facilities." *Clinical Nursing Research,* 2018.

[3] The Minnesota-funded INTERACT® (Interventions to Reduce Acute Care Transfers) quality improvement program focused on managing acute changes in residents' condition. See https://pathway-interact.com/.

[4] The National Nursing Home Quality Improvement Campaign was introduced in 2016 (a follow-on to the Advancing Excellence Campaign, which launched in 2006).

[5] "Nursing Home Quality Initiatives: Questions and Answers." Available at https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/QAPI/Downloads/Nursing-Home-Quality-Initiatives-FAQ.pdf, and https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/QAPI/Downloads/Nursing-Home-Quality-Initiatives-FAQ.pdf.

Though the pilot intervention ended in 2013, it is possible that it resulted in a persistent improvement in outcomes in CareChoice facilities, with limited incremental improvement from PCCC implementation in the treatment group SNFs. Second, it is possible that comparison SNFs might have had less severely ill patients than treatment SNFs (on unmeasured characteristics) or were in areas with more supporting services, resulting in better outcomes for their patients during and after the SNF stay. If not adequately controlled for, such differences in geographic and facility characteristics would have resulted in underestimation of program effects.

PROGRAM SUSTAINABILITY

After its award ended, CareChoice reported that all 10 participating SNFs used internal funding to continue some aspects of PCCC without oversight from the awardee, although no site continued the program in whole. Most sites reported continued use of nursing home decision-support tools and many reported continued follow-up phone calls to patients after discharge and PCCC-established transition planning responsibilities. The awardee expected that the SNFs would not be able to sustain the program in whole without the level of funding generated by the proposed payment model.

CCC's proposed payment model

CareChoice originally proposed to fund PCCC services through FFS reimbursements for each patient who received enhanced discharge planning services. Participating SNFs could use these one-time payments to pay for the salary of a full-time transition coordinator, the cost of the software that guided the planning process, staff training on the model, and any related supplies. Potential payers would have included any health care entity responsible for the outcomes of its patient population, such as accountable care and managed care organizations. The awardee did not propose any adjustments to the fee based on patients' medical complexity, quality measures, or spending benchmarks.

CareChoice spent its one-year no-cost extension period (from September 2017 through May 2018) analyzing data needed to develop the payment model. However, the awardee decided to forgo the originally proposed payment model that relied on FFS payments after its analyses showed that the program did not achieve its intended goals of reducing costs and readmissions. Such results, the awardee anticipated, would make it difficult to negotiate a payment agreement with payers.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

To estimate program impacts, the study used a post-period cross-sectional regression. The regression estimates the treatment–comparison group differences, controlling for beneficiaries' demographic and clinical characteristics, the value of the outcome variable at baseline, and facility-level cost and use, case mix, star rating, and geographic characteristics. The intervention years are beneficiary-specific and defined relative to each beneficiary's enrollment or pseudo-enrollment date. Appendix A of Volume I of this report provides details on the general post-period cross-sectional regression modeling strategy and the standard set of core outcomes used for this evaluation.

The evaluation defined participants as all beneficiaries who were admitted to a participating skilled nursing facility (SNF) from January 1, 2015, through March 19, 2017 and discharged home, as identified from Medicare fee-for-service claims data. Of 4,300 Medicare beneficiaries admitted to participating SNFs during the study intake period, 2,191 (51 percent) were discharged to a community setting, and thus defined as participants. The impact analysis included 41 percent (n = 900) of the 2,191 total participants (Table A.1). The study dropped most of the excluded participants because they did not meet the study's standard claims-based inclusion criteria, with Medicare Advantage enrollment being the primary reason.

Table A.1. Identification of the final sample for impact analysis for CCC

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total number of enrollees		**8,106**
Did not have Medicare claim at a participating SNF	3,716	4,300
Transferred to another SNF facility	1,042	3,258
SNF stay did not result in a discharge to community	1,067	2,191
Did not meet standard claims-based inclusion criteria		
Not enrolled in both Medicare Parts A and B	33	2,158
Enrolled in Medicare Advantage	1,127	1,031
Medicare was not the primary payer	9	1,022
Lacked 90 days of FFS enrollment during baseline period	18	1,004
Did not have claim for hospitalization in the 30 days preceding SNF admission	36	968
Did not match with MDS data	68	900
Final analytic sample		**900**

Sources: Mathematica's analysis of SNF admissions from January 1, 2015, through March 19, 2017, Minimum Dataset (MDS), and information from Medicare claims and enrollment data as of November 2017.

CCC = CareChoice Cooperative; FFS = fee-for-service; MDS = minimum data set; SNF = skilled nursing facility.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. The table displays the weighted means of baseline characteristics for the 900 treatment beneficiaries and the 2,563 matched comparison beneficiaries used in the impact analysis. The table shows the means, adjusted difference in means, the percentage difference, and the standardized difference for each matching variable. The study calculated the standardized difference as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables included demographic characteristics; Medicare entitlement and dual eligibility status; health status (measured by the hierarchical condition category [HCC] score and selected chronic condition indicators); skilled nursing facilities (SNF) admission characteristics; total Medicare expenditures; service use; and facility-level factors. An exact match was required for quarter of SNF admission. The study measured variables over various specified intervals within the 12 months before treatment group enrollment date and comparison group pseudo-enrollment date. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

The table also shows the results of the equivalency-of-means tests. The p-values come from a weighted two-sample t-test, which provides evidence of a statistically significant difference in the means. The equivalence test *p*-values are the greater of the two one-sided weighted t-test p-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the study performed an omnibus test in which the null hypothesis is that the treatment and matched comparison groups are balanced across all linear combinations of the covariates. The study used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

Table B.1. Baseline characteristics of treatment and matched comparison groups for CCC

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	80 (0.36)	80 (0.22)	-0.28 (0.50)	< +/-1	-0.03	0.58	< 0.01
Female, %	67 (1.6)	65 (0.94)	1.5 (2.3)	2.3	0.03	0.51	< 0.01
Married, %	33 (1.6)	33 (0.91)	0.15 (2.2)	< +/-1	0.00	0.95	< 0.01
White, %	92 (0.92)	91 (0.56)	0.42 (1.3)	< +/-1	0.02	0.75	< 0.01
Medicare entitlement and dual eligibility status, %							
Dually eligible for Medicare and Medicaid	14 (1.2)	15 (0.71)	-0.39 (1.7)	-2.7	-0.01	0.82	< 0.01
Original reason for Medicare entitlement: age	83 (1.2)	84 (0.76)	-0.52 (1.7)	< +/-1	-0.01	0.76	< 0.01
Original reason for Medicare entitlement: disability	16 (1.2)	16 (0.75)	0.53 (1.7)	3.2	0.01	0.75	< 0.01
Original reason for Medicare entitlement: ESRD	0.33 (0.19)	0.34 (0.12)	-0.01 (0.27)	-2.2	0.00	0.98	< 0.01
Health status and diagnoses							
HCC score[a]	2.48 (0.05)	2.52 (0.03)	-0.04 (0.07)	-1.5	-0.03	0.58	< 0.01
Acute renal failure, %	25 (1.4)	25 (0.85)	0.08 (2.1)	< +/-1	0.00	0.97	< 0.01
Cardiorespiratory failure and shock, %	19 (1.3)	19 (0.76)	0.65 (1.8)	3.4	0.02	0.72	< 0.01
Chronic obstructive pulmonary disorder, %	19 (1.3)	19 (0.78)	0.17 (1.9)	< +/-1	0.00	0.93	< 0.01
Coagulation defects and other specified hematological disorders, %	14 (1.2)	14 (0.70)	-0.25 (1.6)	-1.8	-0.01	0.88	< 0.01
Congestive heart failure, %	32 (1.6)	34 (0.91)	-1.8 (2.2)	-5.7	-0.04	0.41	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Diabetes with chronic complications, %	19 (1.3)	19 (0.76)	-0.14 (1.8)	< +/-1	0.00	0.94	< 0.01
Major depressive, bipolar, and paranoid disorders, %	17 (1.2)	18 (0.75)	-0.80 (1.8)	-4.7	-0.02	0.65	< 0.01
Morbid obesity, %	10 (1.0)	10 (0.60)	0.38 (1.4)	3.7	0.01	0.79	< 0.01
Septicemia, sepsis, and systemic inflammatory response syndrome or shock, %	14 (1.1)	14 (0.67)	-0.76 (1.7)	-5.6	-0.02	0.64	< 0.01
Specified heart arrhythmias, %	37 (1.6)	38 (0.95)	-0.46 (2.3)	-1.2	-0.01	0.84	< 0.01
Vascular disease, %	24 (1.4)	25 (0.84)	-0.97 (2.0)	-4.1	-0.02	0.63	< 0.01
Admission characteristics							
ADL score	15 (0.16)	15 (0.09)	-0.23 (0.23)	-1.5	-0.05	0.33	< 0.01
CMI ranking: 10 or less, %	4.0 (0.65)	4.0 (0.27)	-0.01 (0.92)	< +/-1	0.00	0.99	< 0.01
CMI ranking: 11 to 30, %	2.4 (0.52)	2.7 (0.32)	-0.26 (0.77)	-11	-0.02	0.74	< 0.01
CMI ranking: 31 to 50, %	38 (1.6)	36 (0.95)	1.8 (2.3)	4.7	0.04	0.42	< 0.01
CMI ranking: 51 or more, %	55 (1.7)	57 (0.97)	-1.5 (2.2)	-2.8	-0.03	0.49	< 0.01
RUG category	190 (3.5)	190 (1.9)	-0.45 (4.9)	< +/-1	0.00	0.93	< 0.01
Medicare expenditures							
Total expenditures	2,564 (74)	2,508 (39)	57 (99)	2.2	0.03	0.57	< 0.01
Total expenditures, 3 months before enrollment	6,972 (209)	6,799 (107)	173 (274)	2.5	0.03	0.53	< 0.01
Total expenditures, 7 days before enrollment	52,780 (1,413)	51,809 (815)	971 (1,996)	1.8	0.02	0.63	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Service utilization							
Total hospitalizations	1,489 (30)	1,468 (16)	21 (42)	1.4	0.02	0.61	< 0.01
Total hospitalizations, 3 months before enrollment	4,556 (57)	4,551 (29)	4.5 (76)	< +/-1	0.00	0.95	< 0.01
Total ED or observation visits	793 (57)	766 (23)	26 (74)	3.3	0.02	0.72	< 0.01
Total ED or observation visits, 3 months before enrollment	1,013 (86)	953 (34)	60 (116)	5.9	0.02	0.60	< 0.01
Total ED or observation visits, 3 months before enrollment[b]	916 (68)	916 (32)	0.00 (98)	< +/-1	0.00	1.00	< 0.01
Primary care visits, any setting	11,770 (277)	11,801 (163)	-31 (405)	< +/-1	0.00	0.94	< 0.01
Primary care visits, any setting, 3 months before enrollment	21,942 (567)	21,917 (319)	25 (805)	< +/-1	0.00	0.97	< 0.01
Length of the most recent hospitalization	6.1 (0.13)	6.2 (0.07)	-0.07 (0.17)	-1.1	-0.02	0.69	< 0.01
Any readmission, %	4.9 (0.72)	4.2 (0.27)	0.67 (1.0)	14	0.03	0.51	< 0.01
ICU or CCU use in prior hospitalization, %	23 (1.4)	23 (0.85)	0.70 (1.9)	3.0	0.02	0.72	< 0.01
Facility-level factors							
Metropolitan area, %	96 (0.65)	96 (0.28)	-0.27 (0.89)	< +/-1	-0.01	0.76	< 0.01
SNF size (number of beds)	121 (1.7)	117 (1.1)	4.2 (2.6)	3.4	0.08	0.11	< 0.01
Overall SNF five-star rating	3.6 (0.04)	3.7 (0.02)	-0.11 (0.05)	-3.1	-0.10	0.02	< 0.01
Facility expenditures	3,193 (19)	3,253 (12)	-60 (28)	-1.9	-0.10	0.04	< 0.01
Facility-level post-discharge ED or observation admissions	0.45 (0.00)	0.45 (0.00)	0.00 (0.00)	< +/-1	-0.05	0.47	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Facility-level post-discharge hospitalizations	0.43 (0.00)	0.43 (0.00)	-0.01 (0.00)	-1.4	-0.07	0.15	< 0.01
Propensity score	0.24 (0.01)	0.23 (0.00)	0.01 (0.01)	3.6	0.05	0.23	< 0.01
Number of beneficiaries	900	2,563					
Omnibus test				Chi-squared statistic	Degrees of freedom	*p*-value	
				681.94	50.00	0.00	

Source: Mathematica's analysis of SNF admissions from January 1, 2015, through March 19, 2017, Minimum Dataset, Nursing Home Compare, and information from Medicare claims and enrollment data as of November 2017.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. p-values come from a weighted two-sample t-test; equivalence test p-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. An exact match was required for quarter of SNF admission.

Higher CMI rankings indicate greater resource use.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

ADL = activities of daily living; CCC = CareChoice Cooperative; CCU = coronary care unit; CHF = congestive heart failure; CMI = case mix index; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; ICU = intensive care unit; RUG = resource utilization group; SE = standard error; SNF = skilled nursing facility.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Table C.1 displays the results from the impact analysis measured separately over intervention Years 1 and 2 for core outcomes and cumulatively for awardee-specific outcomes. The analysis estimated the models over Medicare expenditures; number of services used (per 1,000 beneficiaries); and probability of hospital admissions, skilled nursing facility readmissions, and death within 90 days of discharge. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the CCC intervention on selected Medicare FFS expenditures (dollars PBPM) and use measures for beneficiaries during one- and two-year follow-up periods

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)					
Baseline year	2,530	2,490			
Year 1	2,921	2,756	164 (232)	6.0%	0.48
Year 2	1,516	1,406	109 (147)	7.8%	0.46
Outpatient expenditures ($ PBPM)					
Baseline year	229	272			
Year 1	237	283	-46* (24)	-16%	0.06
Year 2	190	205	-15 (34)	-7.5%	0.65
Hospital stays, per 1,000 beneficiaries					
Baseline year	1,462	1,457			
Year 1	676	663	13 (70)	1.9%	0.86
Year 2	536	494	42 (64)	8.4%	0.51
ED or observation visits, per 1,000 beneficiaries					
Baseline year	790	767			
Year 1	795	1,007	-212*** (79)	-21%	< 0.01
Year 2	630	790	-160* (83)	-20%	0.05
Hospital readmissions, per 1,000 hospital discharges					
Baseline year	15	13			
Year 1	11	11	-0.03 (1.4)	< 1%	0.98
Year 2	16	16	-0.89 (2.7)	-5.4%	0.74
Primary care visits in any setting, per 1,000 beneficiaries					
Baseline year	11,652	11,772			
Year 1	16,335	16,205	130 (866)	< 1%	0.88
Year 2	10,329	10,218	111 (615)	1.1%	0.86
Specialist visits in any setting, per 1,000 beneficiaries					
Baseline year	12,628	11,570			
Year 1	10,507	10,313	193 (774)	1.9%	0.80
Year 2	8,703	8,636	67 (561)	< 1%	0.90

Table C.1 (*continued*)

	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital admissions within 90 days of SNF discharge					
Cumulative	18	18	-0.47 (2.3)	-2.6%	0.84
SNF readmissions within 90 days of SNF discharge					
Cumulative	8.1	9.7	-1.6 (1.7)	-16%	0.35
Death within 90 days of SNF discharge					
Cumulative	4.9	5.7	-0.81 (1.3)	-14%	0.54
Sample sizes					
Number of beneficiaries					
Baseline year	900	2,563			
Year 1	900	2,563			
Year 2	715	1,979			
Number of index discharges from SNFs for readmissions					
Baseline year	366	876			
Year 1	1,403	4,019			
Year 2	316	918			

Sources: Mathematica's analysis of SNF admissions from January 1, 2015, through March 19, 2017, and information from Medicare claims and enrollment data as of November 2017.

Note: Impact estimates for all outcomes represent regression-adjusted treatment–comparison differences based on a post-period cross-sectional regression that controls for beneficiaries' characteristics and the value of the outcome variable at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CCC = CareChoice Cooperative; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SNF = skilled nursing facility.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for CareChoice Cooperative (CCC) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to CCC. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on CMS' four core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for CMS' four core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for CCC led to a Bayesian estimate of the program's impact on total Medicare expenditures of less than -1 percent (an estimated reduction of $6 per beneficiary per month) in the first year and an impact of less than 1 percent (an estimated increase of $8 per beneficiary per month) in the second year.

Table D.1. Comparison of frequentist and Bayesian impact estimates for CCC in the first two years after enrollment

Outcome	Follow-up period	Impact estimate (95 percent interval)		Percentage impacts		
		Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	Year 1	164 (-290, 619)	-5.6 (-188, 162)	1%	6%	> -1%
	Year 2	109 (-179, 398)	8.2 (-83, 95)	2%	8%	< 1%
Hospital admissions	Year 1	13 (-124, 150)	0.18 (-44, 41)	1%	2%	< 1%
	Year 2	42 (-84, 167)	4.0 (-29, 35)	2%	8%	< 1%
ED visits	Year 1	-212 (-368, -56)	-5.3 (-74, 57)	< 1%	-21%	> -1%
	Year 2	-160 (-322, 2.1)	2.0 (-51, 53)	2%	-20%	< 1%
Readmissions	Year 1	0.00 (-0.03, 0.03)	0.00 (-0.01, 0.01)	1%	> -1%	> -1%
	Year 2	-0.01 (-0.06, 0.04)	0.00 (-0.01, 0.01)	2%	-5%	< 1%

Source: Mathematica's analysis of information from the awardee's finder file through March 19, 2017 and Medicare claims and enrollment data as of November 2017. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

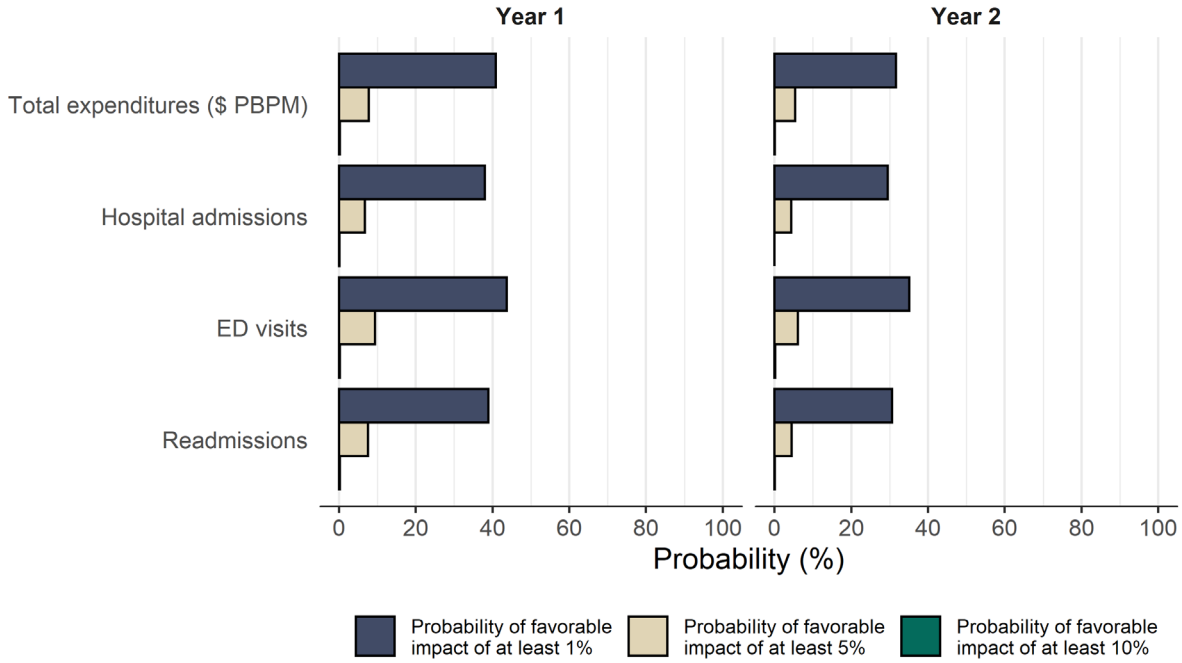
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results are imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that CCC achieved favorable impacts during each of the first two years on the four core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the CCC program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from the awardee's finder file through March 19, 2017 and Medicare claims and enrollment data as of November 2017. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a modest probability—in the range of 40 percent in Year 1 and 30 percent in Year 2—that CCC had a favorable impact of 1 percent or more on total Medicare expenditures, hospital admissions, emergency department visits, and readmissions. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the CCC program did not have a meaningful impact on total expenditures or service utilization.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Catholic Health Initiatives Iowa Corporation

September 2020

Boyd Gilman, Allison Steiner, and Lee-Lee Ellis

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

CATHOLIC HEALTH INITIATIVES IOWA CORPORATION

The Mercy Accountable Care Organization (ACO), which is headquartered in Des Moines, Iowa, and is a division of Catholic Health Initiatives Iowa Corporation (CHIIC), received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Transitioning a Rural Health Network to Value-Based Care program. The goal of the program was to expand population health activities—such as health coaching at primary care clinics, a disease registry throughout the ACO network, and quality improvement projects at hospitals and clinics—to rural, low-income communities in Iowa and Nebraska with high rates of diabetes, obesity, and disability. The target population included rural residents who either had one or more specific chronic conditions (diabetes, hypertension, chronic obstructive pulmonary disease [COPD], and cardiovascular disease) or emergency department (ED) utilization for non-emergent management of chronic diseases. The program was launched in September 2014 and the intervention period covered under HCIA R2 ended in February 2018, after a six-month no-cost extension. Table 1 provides a summary of the key characteristics of the program.

Important issues for understanding the evaluation

- The program aimed to improve health, increase the use of primary care, reduce the use of the ED, and reduce costs for beneficiaries in rural regions of Iowa and Nebraska with certain chronic conditions through health coaching and other population health activities.
- The program represented an expansion of an existing program that urban clinics that were members of the Mercy ACO network had already implemented.
- This impact analysis relied on 1,924 Medicare FFS beneficiaries who had an in-person visit with a health coach and 7,560 comparison beneficiaries with similar demographic and health characteristics who had a primary care visit at CAH-affiliated clinics in the same region that were not members of the Mercy ACO network, and thus ineligible to participate in the program.
- Of the roughly 15,000 Medicare beneficiaries who qualified for the study, only about 3,500 (23 percent) received one or more visits from a health coach. Analyses of the eligible pool suggested that self-selection did not bias the impact estimates based on participants.

Awardee leaders hypothesized that through expanded population health activities, substantial gains could be made in improving population health in poor, rural communities located in the Mercy Health Networks' three geographic regions: Central Iowa (12 hospitals), North Iowa (9 hospitals), and Siouxland (located in western Iowa and eastern Nebraska, with 4 hospitals). Each participating hospital had one or more affiliated primary care clinics where the program's health coaches deliver program services. An innovative feature of the program was that it incorporated critical access hospitals (CAHs) and other rural hospitals into Mercy ACO's existing Medicare shared savings program (MSSP). The goals of the program were to (1) improve population health, as measured by the Centers for Medicare & Medicaid Services' 33 ACO quality measures; (2) increase use of

primary care, decrease use of the ED for non-emergency conditions, and reduce preventable hospitalizations; and (3) reduce the total cost of care for participants.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	CHIIC implemented the program to improve management of chronic conditions among Medicare and Medicaid FFS beneficiaries in rural Iowa and Nebraska and reduce the cost of care for these beneficiaries.
Major innovation	CHIIC expanded existing urban-based population health activities to Medicare and Medicaid FFS beneficiaries in resource-limited, rural communities. In addition, the program was one of the first attempts to incorporate rural hospitals and their clinics into a Medicare ACO.
Program components	• Health coaching over 6 weeks to assess gaps in care, coordinate care among medical providers, connect patients to community resources, and help participants set and achieve health goals • Quality improvement to identify areas in which to improve clinic operations, standardize work flow processes, and reduce costs • Health IT, including use of care management software and a disease registry, to identify patients, manage care, and report outcomes
Target population	Rural residents with one or more chronic conditions (diabetes, hypertension, COPD, and cardiovascular disease) or patients with ED utilization for non-emergent management of chronic diseases.
Participating providers	26 hospitals and 75 affiliated rural clinics participated in the program. Of these, the analysis excluded 3 non-CAHs (and their 16 rural clinics) because the hospitals had already adopted health coaching in their urban clinics. The impact analysis relies on 23 CAHs and their 59 affiliated rural clinics.
Total enrollment	A total of 6,489 patients were enrolled in health coaching from September 2014 through August 2017 (five times the awardee's original enrollment goal of 1,295). Cumulative total individuals ever enrolled include only patients who had an initial face-to-face visit through health coaching services.
Level of engagement	Among all enrollees reported in the awardee encounter database, 36 percent had one visit with a health coach, 31 percent had two or three visits, and 13 percent had four or five visits. Only 20 percent of participants had the recommended six or more visits with a health coach.
Theory of change or theory of action	Health coaches and other medical staff at rural primary care clinics identified patients in rural communities who would benefit from health coaching. After individuals agreed to participate, the nurse health coach performed a range of care management activities, including goal-setting, medication review, depression screenings, cancer screening referrals, and self-management support. The health coaching intervention made patients more accountable for their health and closed gaps in care, which increased the use of primary care, raised vaccination and screening rates, and reduced unnecessary inpatient and ED use. These outputs were, in turn, expected to lead to better health, appropriate health care use, and lower costs.
Award amount	$10,170,496
Effective launch date	• Health coaching component became operational in September 2014 • Impact analysis includes participants enrolled during first three years of program, through August 2017

Table 1 (*continued*)

Program characteristics	Description
Program settings	Primary care clinics affiliated with rural hospitals
Market area	Rural counties in Iowa and eastern Nebraska
Target outcomes	• Improved population health, as measured by CMS's 33 ACO quality measures • Increased use of primary care, less use of the ED for non-emergency conditions, and fewer preventable hospitalizations • Reduced total cost of care for participants
Payment model	The awardee created a new rural MSSP ACO for CAHs and their affiliated clinics in the Mercy Health Network. The shared savings arrangement was supplemented with billable population health activities, such as transitional care management and annual wellness exams, conducted by the nurse health coaches.
Sustainability plans	At the end of the award, all rural hospitals participating in the health coaching program joined the new rural MSSP ACO. All new clinics that join the Mercy ACO in the future will implement the health coaching program. Mercy ACO also continued to invest in the health IT platform to ensure the hospitals, clinics, and the ACO can share clinical data.

ACO = accountable care organization; CAH = critical access hospital; CHIIC = Catholic Health Initiatives Iowa Corp.; CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disease; ED = emergency department; FFS = fee-for-service; IT = information technology; MSSP = Medicare shared savings plan.

The impact analysis presented in this report included 1,924 Medicare fee-for-service (FFS) beneficiaries who had an in-person meeting with a health coach (which the awardee considered a requirement of enrollment) and met the other claims-based study inclusion criteria. The study identified a propensity score matched-comparison group from Medicare enrollment and claims data. The comparison group included 7,560 Medicare FFS beneficiaries with similar demographic and health characteristics who had a primary care visit at CAH-affiliated clinics in the same rural counties that were not members of the Mercy ACO network and thus ineligible to participate in the program. Table 2 summarizes the key features of the evaluation.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group.
Intervention group for evaluation	The impact analysis relied on 1,924 Medicare FFS beneficiaries who enrolled in health coaching from September 2014 to August 2017. The study sample did not include an estimated 1,372 patients from the Siouxland Region who enrolled after October 2016 when hospitals in that region changed their care management software and were no longer able to link with the awardee's reporting system. It also excluded 2,634 patients who either were not enrolled in Medicare or could not be identified in the Medicare enrollment database through names-based matching, and 559 beneficiaries who did not meet the claims-based eligibility criteria of the study, such as being enrolled in Parts A and B

Table 2 (*continued*)

Features	Description
Comparison group	The comparison group included 7,560 Medicare FFS beneficiaries with similar demographic and health characteristics who had a primary care visit at CAH-affiliated clinics in the same rural counties that were not members of the Mercy ACO network and thus ineligible to participate in the program.
Limitations	If participants differed from eligible nonparticipants in ways not captured in Medicare administrative files and claims, the impact estimates might be biased. The low participation rate (about 23 percent) would have made it difficult to identify impacts if measured over all eligible beneficiaries.

ACO = accountable care organization; CAH = critical access hospital; and FFS = fee for service.

PROGRAM DESIGN AND ADAPTATION

The Transitioning a Rural Health Network to Value-based Care program service delivery model had three components: (1) health coaching, (2) continuous quality improvement, and (3) health information technology (health IT).[1]

Health coaching

The health coaching component provided short-term, intensive care management for adults with one or more chronic diseases, including diabetes, hypertension, COPD, and cardiovascular disease, or frequent ED visits. The component required that health coaches be registered nurses. Interactions with a nurse health coach were intended to take place at least once a week for six weeks. Coaching sessions preferably occurred in person, although they could have taken place over the phone, especially after the initial face-to-face meeting. Health coaches provided a range of care management services, including (1) reviewing medical conditions and monitoring test values, (2) reviewing medication lists, (3) providing self-management support and goal-setting, (4) screening for depression and use of tobacco, (5) closing gaps in medical care, and (6) coordinating follow-up care. Health coaches described these offerings more as guides than as a formal protocol, and their regional managers encouraged the coaches to tailor services to the needs of their patients. The frequency and duration of meetings varied depending on participants' level of engagement and their progress meeting their health coaching goals.

Quality improvement

A second practice-level component focused on continuous quality improvement and each group of hospitals and their affiliated clinics implemented it. It used Lean Process Improvement principles to help reduce costs and optimize operations. Quality improvement activities varied based on the needs and priorities of each rural hospital and its affiliated clinics.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Health IT

Finally, the program included a health IT component. The Mercy ACO began to create an online platform to connect its data feeds with participating clinics and hospitals, host a disease registry to track billing and clinical information for the ACO's patient population, and provide data to hospital and clinic administrators for quality improvement activities.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee quickly ramped up the operations for its health coaching component, reaching its three-year staffing goals by the end of the first year. The awardee also surpassed its initial enrollment goal for health coaching in the first year and raised its enrollment target several times throughout the award period. Early implementation was due in part to a successful health coaching initiative in Mercy's urban clinics that began in 2012, which the ACO slowly began to introduce in its rural clinics in 2013, one year before the start of the HCIA. Program staff also reported that the small size of the participating rural clinics made it easier to promote the program's visibility, facilitate referrals, and integrate its services into the clinical workflow. Staff attributed the successful implementation in the first year to a strong commitment to population health among Mercy ACO leadership as well.

However, the awardee faced major challenges implementing the two practice-level components of the program. For the quality improvement component, the awardee faced challenges with staff turnover and the availability of reliable data for identifying quality improvement activities. Furthermore, local priorities determined these activities, so they varied across participating sites. Regarding the health IT component, the awardee was unable to connect all sites' emergency medical record systems to the disease registry due to the diversity of platforms across the network. Awardee leaders reported that health IT development was much more challenging than they anticipated. The independent impact of each intervention component on outcomes cannot be measured.

Implications of program implementation for detecting impacts

- Because the new health coaching program represented an expansion of an existing initiative in urban practices and had been slowly introduced in rural clinics before the HCIA, participating sites as a whole reached their implementation goals early in the first year.
- The two practice-level components of the award (practice transformation and health IT), might have influenced the impact of the health coaching intervention, but the concurrent implementation of the three components makes it impossible to assess their independent effects on outcomes.

ESTIMATING PROGRAM IMPACTS

Study sample

The treatment group sample for the impact analysis was based on 1,924 Medicare FFS beneficiaries who had at least one face-to-face encounter with a health coach from September 1, 2014, through August 31, 2017 (representing 30 percent of the 6,489 total enrollment during that period) The treatment group excluded an estimated 1,372 patients from the Siouxland Region who enrolled after October 2016 when hospitals in that region changed their care management software and could no longer link with the awardee's reporting system. It also excluded 2,634 patients who either were not enrolled in Medicare or could not be identified in the Medicare enrollment database through names-based matching, and 559 beneficiaries who did not meet the other standard study inclusion criteria, such as being enrolled in Parts A and B, having Medicare as a primary payer, and having at least 90 days of Medicare enrollment in the baseline period. (Appendix A, Table A.1 describes the identification of the treatment group for the impact analysis).

The study sample also included 7,560 matched comparison beneficiaries who had a primary care visit from September 2014 through August 2017 at one of the 27 CAHs in the same rural counties but were not members of the Mercy ACO network and thus unable to participate in the program. Appendix A provides additional detail on the identification of the comparison group for this study.

Enrolling participants

Health coaches identified patients for coaching in several ways. They identified eligible participants by conducting chart reviews before scheduled visits, generating lists of patients with targeted chronic conditions from electronic medical record registries, and reviewing hospitals' administrative lists of patients with ED visits or hospitalizations for management of chronic diseases. The enrollment protocol did not provide specific ED or inpatient use criteria. In addition, primary care providers and other clinic staff referred patients they believed would benefit from health coaching services, even if they did not meet the formal eligibility criteria. Awardee leaders acknowledged that, although all clinics tried to enroll patients with hypertension and diabetes, some practices focused on patients with other needs (such as smoking cessation and weight loss) based on the referring providers' interests, community health needs, and billing opportunities. After patients were identified as eligible or directly referred into the program, they had an initial in-person meeting with a health coach, who introduced the goals of the program and expectations of participation and confirmed the patient's willingness to participate.

Characteristics of treatment and comparison group beneficiaries

A comparison of treatment and comparison group characteristics at baseline confirmed that the two groups were well balanced (Table 3). The average age of treatment and comparison group

beneficiaries during the baseline year was 73 years. Hypertension was by far the most common target condition, accounting for roughly 80 percent of both groups, followed by diabetes with or without chronic conditions at about 44 percent. Slightly more than 10 percent in both groups did not have any of the four target conditions; presumably, the program enrolled them because their physicians believed the health coaching program could help them address other health goals, such as smoking cessation and weight loss. Mean predicted expenditures for both groups (measured by their hierarchical condition category [HCC] scores) were nearly 30 percent higher than the average for Medicare FFS beneficiaries nationally. Appendix B provides the full balance results measured during the 12 months before enrollment.

Beneficiaries who enrolled in the health coaching program during the first few months of the program appeared to be sicker on average than those who enrolled later (Table 3). Beneficiaries who enrolled within the first nine months of the program start date were more likely to have been originally eligible for Medicare because of disability than later enrollees. They were also more likely to have had at least one of the four targeted conditions and slightly higher HCC scores at baseline. Later enrollees were twice as likely to have none of the target conditions. Consistent with these differences, early enrollees had significantly higher Medicare expenditures and were much more likely to have used inpatient and ED services during the year before they enrolled in the health coaching intervention than later enrollees.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

	Full group		Treatment group only	
Measure	Treatment (N = 1,924)	Comparison (N = 7,560)	Enrolled in first 9 months (N = 179)	Enrolled after first 9 months (N = 1,745)
Demographics				
Age at enrollment, years	73	73	72	74
Age group, %				
Younger than 65	11	11	14	11
65 to 74	42	44	45	42
75 to 84	32	30	29	32
85 and older	14	14	11	15
Male, %	41	40	49	40
White, %	98	98	98	98
Original reason for Medicare eligibility, %				
Old age and survivor's insurance	81	81	79	81
Disability insurance benefits[a]	19	19	21	19
Medicare/Medicaid dual status, %	15	15	13	16
Target conditions, %				
COPD	16	16	19	16
Diabetes without other chronic conditions	24	24	43	22
Diabetes with other chronic conditions	20	19	20	20

Table 3 (*continued*)

Measure	Full group		Treatment group only	
	Treatment (N = 1,924)	Comparison (N = 7,560)	Enrolled in first 9 months (N = 179)	Enrolled after first 9 months (N = 1,745)
Hypertension	82	81	86	81
Vascular disease	18	17	19	18
None of the target conditions	11	12	5	11
HCC score[b]				
Mean	1.29	1.28	1.36	1.29
25th percentile	0.57	0.55	0.61	0.56
Median	0.94	0.94	0.97	0.94
75th percentile	1.64	1.58	1.69	1.64
Service use and expenditures during the year before enrollment				
Any hospitalizations, %	24	23	36	23
Any outpatient ED visits, %	36	36	45	35
Number of hospital admissions (per 1,000)	346	354	537	327
Number of outpatient ED visits (per 1,000)	607	617	820	585
Total Medicare expenditures ($ PBPM)	956	969	1,249	926

Source: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data from September 2013 to November 2017, as of November 30, 2017.

Note: The study defined the baseline year as the 365 days before each beneficiary's enrollment date. The study defined the enrollment date as the date of a participant's first face-to-face encounter with a health coach. The study measured all beneficiary characteristics during or as of the end of the baseline year. None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level, two-tailed test. The study used exact matching on quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, state of residence, and rural health clinic enrollment status. Appendix B presents the full balance results.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the statistics for comparison beneficiaries are weighted to reflect the number of times a comparison beneficiary is matched to a treatment beneficiary.

[a] Includes participants with both a disability and ESRD.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disorder; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact estimates are based on a difference-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables will capture differences on unobserved variables as well. The primary outcomes are

total Medicare spending, number of hospital admissions, and number of ED visits. Secondary outcomes include number of primary care and specialty care visits.

The pre-enrollment period is defined as the year before each participant's enrollment date and the post-enrollment period is defined as the two years after. The enrollment date is defined as the date of a participant's first face-to-face encounter with a health coach. The study team assigned a pseudo-enrollment date to each comparison beneficiary, based on the date of a matched primary care visit. The team obtained separate estimates for participants who enrolled within the first nine months of the program start date (September 1, 2014–May 31, 2015) and after the first nine months (June 1, 2015–August 31, 2017). Appendix A describes the statistical models and outcomes used to estimate the effects of the program, as well as the identification of the final analytic sample.

IMPACT RESULTS

The estimated effects of the health coaching intervention measured over all program participants on Medicare spending and the use of inpatient and ED services were small and not statistically significant (Table 4). However, the effects were larger among the small set of participants who enrolled within the first nine months of the program start date (about 10 percent of the total). The estimated reduction in total spending for early enrollees in Year 1 was relatively large (**24 percent**) and statistically significant. An estimated reduction in inpatient stays for this subgroup of early enrollees appears to have driven this estimated effect; to a lesser extent, the estimated reduction in ED visits also contributed. These estimates were large but not statistically significant. There was also a substantial and statistically significant estimated increase in primary care visits (with the effect concentrated in the first year of enrollment) and a decrease in specialty care visits (with the effect concentrated in the second year of enrollment) over all participants. Appendix C presents the full results of the impact analysis. Appendix D shows the results from a Bayesian analysis.

Table 4. Estimated impact of health coaching on selected outcomes

	Full group		Enrolled during first 9 months		Enrolled after first 9 months	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Expenditures ($ PBPM)						
Impact ($)	19	-13	-327	-285	56	20
Percentage impact	1.9%	-1.3%	-24%	-22%	5.2%	1.9%
p-value	0.78	0.86	0.09	0.19	0.44	0.80
Number of hospitalizations, per 1,000 beneficiaries						
Impact (rate)	-3.7	-26%	-100	-116	6.4	-14
Percentage impact	-1.0%	-6.6%	-21%	-24%	1.9%	-4.3%
p-value	0.89	0.41	0.31	0.22	0.83	0.68

Table 4 (*continued*)

	Full group		Enrolled during first 9 months		Enrolled after first 9 months	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Number of ED or observation visits, per 1,000 beneficiaries						
Impact (rate)	15	-50	-154	-101	33	-46
Percentage impact	1.9%	5.4%	-16%	-10%	5.4%	-7.2%
p-value	0.70	0.29	0.26	0.53	0.42	0.35
Number of specialty care visits, per 1,000 beneficiaries						
Impact (rate)	-302	-821	36	-1,012	-337	-782
Percentage impact	-3.7%	-10%	< 1%	-12%	-4.4%	-10%
p-value	0.22	0.01	0.96	0.35	0.20	0.01
Number of primary care visits, per 1,000 beneficiaries						
Impact (rate)	664	287	515	-66	679	339
Percentage impact	7.6%	3.1%	5.8%	< 1%	9.2%	4.3%
p-value	< 0.01	0.25	0.41	0.93	< 0.01	0.20
Sample size						
Treatment	1,924	1,924	179	179	1,745	1,745
Comparison	7,560	7,560	671	671	6,889	6,889

Source: Mathematica's analysis of information from the awardee's program encounter database from September 1, 2014, through August 31, 2017, and Medicare claims and enrollment data from September 2013 through February 2018, as of March 13, 2019.

Notes: Impact estimates are based on the regression-adjusted difference between the randomized treatment and control group beneficiaries. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the predicted treatment group mean in the post period minus the impact estimate). Appendix C presents full impact estimates. Appendix D shows the results from a Bayesian analysis.

*Significantly different from zero at the .10 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month.

Two factors together might explain why the Mercy ACO's rural health coaching intervention achieved favorable effects on total expenditures and, possibly, on inpatient and ED use among the early group of enrollees but not among later enrollees. First, as previously mentioned, early enrollees appeared to be at higher risk of hospitalizations and ED visits, and thus potentially more likely to achieve improvements in these outcomes than later enrollees. The intervention encouraged clinicians to refer patients with chronic conditions who they believed could benefit from health coaching, so the clinicians might have referred patients with the greatest needs and highest levels of inpatient and ED use first. Over time, chart reviews and disease registries identified more patients. Thus, although they still met the inclusion criteria, these patients might not have had such a high level of need as the initial cohort of patients. Program administrators also reported that, over time, frontline staff began to expand the enrollment criteria, enrolling not only patients with chronic conditions, but also patients with other health goals (such as weight loss and smoking cessation) or those who only had risk factors for chronic conditions. The proportion of participants who enrolled later in the program who had none of the targeted conditions was twice as high as the proportion who enrolled in the first nine months without any

of the conditions. Expanding the eligibility criteria to include risk factors and health goals unrelated to chronic care management likely weakened the program's effects on outcomes in the second and third years of the program.

Main findings from impact evaluation

- The health coaching intervention reduced total Medicare expenditures among the small group of beneficiaries who enrolled within the first nine months of the program start date. A reduction in inpatient stays drove that effect as, to a lesser extent, did a reduction in ED visits.
- Early enrollees had more chronic conditions, used more health care services, and had higher medical expenditures during the year before they enrolled in the program, and thus might have had more to gain from health coaching services, than later enrollees.
- The lower health care needs among the larger group of later enrollees, and the shift in the allocation of nurse health coaching hours, suggest that health coaching services could have a limited ability to reduce expenditures in small rural clinics.

Program administrators provided further evidence that the health care needs of enrollees began to decline after the first program year, and thus help explain why program effects might be stronger among early enrollees. Program administrators reported that, over time, the nurse health coaches began spending less time on health coaching activities and more time providing services that are billable under Medicare FFS, such as transitional care management to beneficiaries recently discharged from a hospital and annual wellness visits. Reflecting the observations of program administrators, program data provided by the awardee show that the average number of visits per participant declined over time. Among participants who enrolled in the first year of the program, 36 percent received the recommended six visits with a health coach, compared with only 16 percent in the second year, and 13 percent in the third year. Health coaches might have begun spending less time on health coaching (and more time on billable services) because they realized they did not need the full six weeks to give participants the skills and information they needed to reach their goals, participants did not want to complete the full six weeks, or there was a general lack of need for health coaching services in small rural clinics.

Second, as previously noted, the program represented an expansion of a similar program that the Mercy ACO had already been operating in its urban clinics for several years and had begun to roll out in some of its rural clinics before the HCIA. Program administrators reported that this prior implementation experience helped them address many of the common start-up challenges before the program began and achieve many of their three-year implementation goals early during the first year of the program. As a result, many participating clinics could provide well-developed interventions soon after the program start date. Although programs typically take a while to have effects, this one might have had impacts for those most likely to benefit right away, because of the prior implementation.

CONCLUSION

Overall, the program had no estimated effect on the primary outcomes. Those who enrolled in the first nine months (only about 10 percent of the total) seemed to experience a favorable effect and appeared to be at somewhat higher risk than later enrollees based on HCC scores and baseline service use. This finding suggests that the program might be effective in lowering costs if appropriately targeted to patients who can benefit the most from the services. However, the number of patients in small rural clinics who can benefit from short-term health coaching might not be enough to warrant hiring a full-time health coach. The participating rural hospitals were able to support a full-time health coaching position only by assigning health coaches to more than one clinic and by requiring health coaches to be registered nurses so that they could provide other billable services under Medicare, such as transitional care management, chronic care management, and annual wellness visits, as their time allowed.

Limitations of evaluation

The analysis has several limitations. First, program impacts were calculated over participants only. If participants differed from eligible nonparticipants in ways not captured in Medicare administrative files and claims, the results of the study cannot be generalized to the full target population. The low participation rate (about 23 percent) would make it difficult to identify impacts if measured over all eligible beneficiaries. Second, the impact analysis does not include all participants from hospitals that changed their care management software half-way through the program and thus were unable to link with the awardee's reporting system. In addition, because the awardee was unable to provide beneficiary identifiers, the study sample includes only those who could be linked to the Medicare enrollment file by name, gender, and date of birth. Differences in the health needs of included versus excluded participants could also bias the results. Third, lack of information on the specific activities that health coaches conducted to improve care management and lack of standardization in the health coaching protocol make it difficult to identify the program features that are most important for achieving favorable outcomes and to replicate the intervention in other settings. Finally, as noted, the awardee initiated two practice-level components designed to support the health coaching program. The independent impact of these intervention components on outcomes could not be estimated.

PROGRAM SUSTAINABILITY

After its award ended in February 2018, CHIIC continued its HCIA R2-funded program through the Mercy Rural MSSP (a rural ACO created during the award period). Hospitals that sign a contract with the ACO must hire a health coach to serve their affiliated clinics for the three-year MSSP contract. The hospitals also must pay the ACO an annual fee, which the awardee uses to continue supporting sites the way it did during the award, such as by employing rural market managers who oversee the health coaches and offering access to health IT platforms that support a disease registry and care management activities.

CHIIC had hoped that implementing sites would be able to pay for program costs from their shared savings with the rural ACO, as well as by billing for covered services conducted by the health coaches, such as annual wellness visits and transitional care management. However, because Medicare's cost-based reimbursement system for CAHs provides a disincentive for hospitals to reduce their average costs, awardee leaders acknowledged that the rural ACO is unlikely to generate shared savings. The awardee hopes that participation in the rural ACO will nonetheless continue and increase due to the other benefits of the program, such as having access to a disease registry and care management software; receiving education on how to increase revenue through FFS billing for population health management activities; and receiving technical assistance with provider credentialing, compliance, and interpretation of CMS rules and regulations.

CHIIC's proposed payment model

CHIIC proposed a payment model that relied on shared savings with its rural ACO, supplemented by FFS billing for covered services that health coaches could conduct, such as transitional care management, advance care planning, and annual wellness visits. To participate in the ACO, hospitals and their affiliated clinics must hire a health coach and pay the ACO an annual fee. The fee helps pay for administrative support from the ACO and provides access to its health IT system.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach with beneficiary-level fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or emergency department (ED) visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiary characteristics and whether the beneficiary had any hospital stay or any ED visit during the baseline period. The intervention years are beneficiary-specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of core outcomes used for this evaluation.

The impact analysis included only 30 percent of total participants, according to awardee data (Table A.1). Participants were defined by having at least one in-person visit with a health coach from September 2014 through August 2017, as reported in the awardee's final encounter database. Most of the excluded participants were dropped from the study because they were not Medicare beneficiaries or could not be found in the Medicare enrollment database through names-based matching (41 percent). Another estimated 21 percent were not included because they enrolled after hospitals in one region switched to a new care management system and could no longer link with the awardee's reporting system. The remaining 9 percent of enrollees were dropped because they did not meet the study's standard claims-based inclusion criteria.

Table A.1. Identification of final sample for impact analysis for CHIIC

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through August 31, 2017		**6,489**
Estimated number of participants from hospitals in nonreporting region after October 2016[a]	1,372	5,117
Participants not enrolled in Medicare or not identified in Medicare enrollment files[b]	2,634	2,483
Participants who did not meet the standard claims-based inclusion criteria		
Not enrolled in both Medicare Parts A and B	259	2,224
Enrolled in Medicare Advantage	210	2,014
Medicare was not the primary payer	24	2,990
Lacked 90 days of FFS enrollment during baseline period	66	1,924
Final analytic sample		**1,924**

Source: Mathematica's analysis of information from the awardee's program encounter database from September 1, 2014, through August 31, 2017, and Medicare claims and enrollment data from September 2013 through February 2018, as of March 13, 2019.

[a] The study sample excluded an estimated 1,000 patients from the Siouxland Region who enrolled after October 2016 when hospitals in that region changed their care management software and could no longer link with the awardee's reporting system.

[b] The awardee did not provide patient identifiers, so Medicare beneficiaries were identified through matching a participant's name, gender, and date of birth with information available from the Medicare enrollment database.

CHIIC = Catholic Health Initiatives Iowa Corp.; FFS = fee-for-service.

The study sample also included 7,560 matched comparison beneficiaries who had a primary care visit at a CAH in the same region that was not part of the Mercy ACO network and thus unable to participate in the program. The multi-step process for selecting the comparison group included (1) identifying the 27 CAHs in the same rural counties as the program but were not a member of the Mercy ACO network; (2) identifying the billing IDs for the 107 clinics affiliated with those CAHs; (3) identifying all FFS Medicare beneficiaries with a professional or outpatient claim with a primary care CPT code from one of the comparison clinics from September 2014 through August 2017 (restricted to providers with a primary care taxonomy code); (4) assigning a pseudo-enrollment date for these beneficiaries based on the date of their first primary care visit within a given follow-up quarter; and (5) excluding beneficiaries who did not meet the standard study inclusion criteria of begin 18 or older, not having end stage renal disease, not being enrolled in Medicare Advantage, being enrolled in both Medicare Parts A and B, having Medicare as their primary payer, and having at least 90 days of Medicare enrollment in the baseline period. The comparison group was also restricted to beneficiaries who were resident of Iowa or Nebraska.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. The table displays the weighted means of baseline characteristics for the 1,924 treatment beneficiaries and the 7,560 matched comparison beneficiaries used in the impact analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable. The standardized difference was calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables included demographic characteristics (age, gender, and race); state of residency; Medicare entitlement and dual eligibility status; enrollment in a rural health center; health status (as measured by the hierarchical condition category [HCC] score and selected chronic condition indicators; Medicare expenditures in total and by type of service; and service use. An exact match was required for the quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, state of residence, and rural health clinic enrollment status. The variables are measured over various specified intervals within the 12 months before enrollment in the intervention.

The table also shows the results of the equivalency-of-means tests. The *p*-values come from a weighted two-sample t-test, which provides evidence of a statistically significant difference in the means. The equivalence test *p*-values are the greater of the two one-sided weighted t-test *p*-values equivalence test, which assesses whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, an omnibus test was performed in which the null hypothesis is that the treatment and matched comparison groups are balanced across all linear combinations of the covariates. The results are used to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

Table B.1. Baseline characteristics of treatment and matched comparison groups for CHIIC

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	73 (0.24)	73 (0.13)	0.27 (0.36)	< +/-1	0.03	0.45	< 0.01
Female, %	59 (1.1)	60 (0.56)	-1.3 (1.6)	-2.2	-0.03	0.42	< 0.01
Age: Younger than 65, %	11 (0.72)	11 (0.36)	0.04 (1.0)	< +/-1	0.00	0.97	< 0.01
Age: 65 to 74, %	43 (1.1)	45 (0.57)	-1.4 (1.6)	-3.2	-0.03	0.39	< 0.01
Age: 75 to 84, %	31 (1.1)	30 (0.53	1.1 (1.5)	3.4	0.02	0.47	< 0.01
Age: 85 and older, %	14 (0.80)	14 (0.39)	0.30 (1.1)	2.1	0.01	0.78	< 0.01
White, %	98 (0.34)	98 (0.16)	0.26 (0.50)	< +/-1	0.02	0.61	< 0.01
Black, %	0.26 (0.12)	0.56 (0.06)	-0.30 (0.21)	-117	-0.06	0.15	< 0.01
Hispanic, %	0.16 (0.09)	0.21 (0.05)	-0.05 (0.14)	-33	-0.01	0.70	< 0.01
Other, %	1.1 (0.24)	0.96 (0.09)	0.13 (0.32)	12	0.01	0.69	< 0.01
Unknown, %	0.73 (0.19)	0.76 (0.10)	-0.03 (0.29)	-4.4	0.00	0.91	< 0.01
Medicare entitlement and dual eligibility status, %							
Original reason for Medicare entitlement: age	81 (0.89)	81 (0.44)	0.18 (1.3)	< +/-1	0.00	0.89	< 0.01
Original reason for Medicare entitlement: disability	19 (0.89)	19 (0.44)	-0.18 (1.3)	< +/-1	0.00	0.89	< 0.01
Dually eligible for Medicare and Medicaid	15 (0.82)	15 (0.41)	-0.04 (1.2)	< +/-1	0.00	0.97	< 0.01
Health status and diagnoses							
HCC score[b]	1.28 (0.03)	1.27 (0.01)	0.01 (0.04)	1.1	0.01	0.69	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Vascular disease, %	18 (0.87)	17 (0.42)	0.94 (1.3)	5.3	0.02	0.46	< 0.01
COPD, %	16 (0.84)	16 (0.41)	-0.11 (1.2)	< +/-1	0.00	0.93	< 0.01
Diabetes with chronic complications, %	20 (0.91)	19 (0.44)	1.3 (1.2)	6.3	0.03	0.30	< 0.01
Diabetes without chronic complications, %	24 (0.98)	24 (0.48)	0.33 (1.4)	1.4	0.01	0.82	< 0.01
CHF, %	16 (0.83)	15 (0.39)	0.50 (1.2)	3.2	0.01	0.67	< 0.01
Hypertension, %	81 (0.89)	80 (0.47)	1.1 (1.2)	1.3	0.03	0.40	< 0.01
Medicare expenditures							
Total expenditures, 12 months before enrollment	936 (40)	961 (18)	-25 (60)	-2.7	-0.01	0.68	< 0.01
Total expenditures, 3 months before enrollment	1,373 (79)	1,436 (36)	-63 (122)	-4.6	-0.02	0.61	< 0.01
Total expenditures, 7 days before enrollment	3,579 (321)	3,111 (66)	46 (394)	13	0.03	0.24	< 0.01
Total expenditures, 7 days before enrollment[a]	2,826 (162)	2,687 (43)	138 (225)	4.9	0.02	0.54	< 0.01
Total expenditures, day of enrollment	1,007 (215)	885 (64)	122 (271)	12	0.01	0.65	< 0.01
Outpatient expenditures, 12 months before enrollment	301 (10)	292 (4.0)	8.7 (14)	2.9	0.02	0.54	< 0.01
Outpatient expenditures, 7 days before enrollment	1,342 (50)	1,227 (14)	114 (70)	8.5	0.05	0.10	< 0.01
Physician service expenditures, 12 months before enrollment	154 (6.3)	156 (2.9)	-2.2 (9.1)	-1.5	-0.01	0.80	< 0.01
Physician service expenditures, 7 days before enrollment	256 (22)	305 (6.8)	-49 (34)	-19	-0.05	0.15	< 0.01
Physician service expenditures, 7 days before enrollment[a]	199 (9.0)	213 (2.9)	-14 (13)	-6.8	-0.03	0.28	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Service use							
Total hospitalizations, 12 months before enrollment	342 (17)	352 (7.7)	-10 (25)	-3.0	-0.01	0.69	< 0.01
Total hospitalizations, 3 months before enrollment	599 (38)	634 (16)	-35 (56)	-5.9	-0.02	0.53	< 0.01
Total ED or observation visits, 12 months before enrollment	720 (38)	699 (14)	21 (50)	2.9	0.01	0.67	< 0.01
Total ED or observation visits, 3 months before enrollment	1,037 (65)	1,005 (25)	32 (85)	3.1	0.01	0.71	< 0.01
Total ED or observation visits, 12 months before enrollment[a]	649 (26)	640 (12)	9.7 (36)	1.5	0.01	0.79	< 0.01
Total ED or observation visits, 3 months before enrollment[a]	971 (56)	954 (23)	16 (75)	1.7	0.01	0.83	< 0.01
Primary care visits, any setting, 12 months before enrollment	7,991 (169)	7,906 (71)	86 (237)	1.1	0.01	0.72	< 0.01
Primary care visits, any setting, 3 months before enrollment	11,143 (245)	11,303 (108)	-160 (352)	-1.4	-0.01	0.65	< 0.01
Any inpatient stay, 30 days before enrollment, %	9.5 (0.67)	9.7 (0.25)	-0.29 (0.95)	-3.1	-0.01	0.76	< 0.01
Any inpatient stay, 2 days before enrollment, %	1.5 (0.27)	1.6 (0.07)	-0.17 (0.41)	-11	-0.01	0.68	< 0.01
Any inpatient services, day of enrollment, %	0.05 (0.05)	0.00 (0.00)	0.05 (0.05)	100	0.02	0.31	< 0.01
Any outpatient ED or observation visit, 30 days before enrollment, %	11 (0.72)	12 (0.30)	-0.17 (1.0)	-1.5	-0.01	0.87	< 0.01
Any outpatient ED or observation visit, 2 days before enrollment, %	2.1 (0.33)	2.0 (0.10)	0.06 (0.47)	2.8	0.00	0.90	< 0.01
Any outpatient ED or observation visit, day of enrollment, %	0.26 (0.12)	0.18 (0.05)	0.08 (0.15)	31	0.02	0.60	< 0.01
Number of beneficiaries	1,924	7,560					
Omnibus test				Chi-squared statistic 70.37	Degrees of freedom 42.00	P-value 0.00	

Table B.1 (*continued*)

Source: Mathematica's analysis of Medicare claims and enrollment data from September 2013 to November 2017, as of November 2017.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include the quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, state of residence, and rural health clinic enrollment status.

[a] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CHF = congestive heart failure; CHIIC = Catholic Health Initiatives Iowa Corp.; COPD = chronic obstructive pulmonary disease; ED = emergency department; HCC = hierarchical condition category; SE = standard error.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Tables C.1 and C.2 display the results from the impact analysis. Table C.1 shows the impact estimates for the full study population, measured separately over intervention Years 1 and 2. Table C.2 shows similar results for the subgroup of 179 (9 percent) treatment beneficiaries who enrolled within the first nine months of the program start date versus the 1,745 (91 percent) of treatment beneficiaries who enrolled after the first nine months of the launch date. The models were estimated over Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. Impact estimates that are statistically different from zero at the .10, .05, and .01 levels, using a two-tailed test, are indicated with one, two, or three asterisks, respectively.

Table C.1. Estimated impact of the CHIIC intervention on selected Medicare FFS expenditures (dollars PBPM) and utilization measures during 1- and 2-year follow-up periods

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)[b]					
Baseline year	956	970			
Year 1	1,146	1,141	19 (68)	1.8%	0.78
Year 2	1,211	1,238	-13 (75)	-1.2%	0.86
Acute inpatient expenditures ($ PBPM)					
Baseline year	313	307			
Year 1	350	361	-18 (41)	-5.2%	0.66
Year 2	370	383	-20 (40)	-6.0%	0.63
Hospital outpatient expenditures ($ PBPM)					
Baseline year	308	296			
Year 1	370	351	6.4 (16)	1.8%	0.69
Year 2	392	401	-21 (21)	-5.6%	0.30
Professional Part B expenditures ($ PBPM)					
Baseline year	156	158			
Year 1	186	175	12 (10)	7.3%	0.21
Year 2	200	189	13 (12)	7.3%	0.28
Primary care visits in all settings, per 1,000 beneficiaries					
Baseline year	8,123	7,987			
Year 1	8,310	7,510	664*** (202)	8.8%	< 0.01
Year 2	8,566	8,144	287 (247)	3.6%	0.25
Specialist visits in all settings, per 1,000 beneficiaries					
Baseline year	6,996	6,769			
Year 1	7,440	7,515	-302 (249)	-3.9%	0.23

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Year 2	7,257	7,852	-821*** (293)	-10%	< 0.01
Hospital stays, per 1,000 beneficiaries[b]					
Baseline year	346	354			
Year 1	365	377	-3.7 (28)	-1.1%	0.89
Year 2	369	403	-26 (31)	-7.3%	0.41
Percentage with a 30-day readmission, among all discharges					
Baseline year	10	13			
Year 1	16	20	2.4 (2.3)	18%	0.30
Year 2	12	17	1.2 (2.5)	11%	0.64
ED or observation visits, per 1,000 beneficiaries					
Baseline year	607	617			
Year 1	661	656	15 (39)	2.3%	0.70
Year 2	652	712	-50 (47)	-7.3%	0.29
Percentage of beneficiaries with any hospital admission in a time period[c]					
Baseline year	24	23			
Year 1	22	21	0.70 (4.1)	3.3%	0.86
Year 2	21	22	-0.38 (4.1)	-1.8%	0.93
Percentage of beneficiaries with any ED or observation visits in a time period[c]					
Baseline year	36	36			
Year 1	34	33	0.73 (5.7)	2.2%	0.90
Year 2	34	36	-1.3 (5.7)	-3.6%	0.82
Sample sizes					
Number of beneficiaries	1,924	7,560			

Source: Mathematica's analysis of information from the awardee's program encounter database from September 2014 through August 2017 and Medicare claims and enrollment data from September 2013 through February 2018, as of March 13, 2019.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiary characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary-specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

[c] Baseline values are the proportion of beneficiaries with more than 2 ED or observation visits in the baseline year

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CHIIC = Catholic Health Initiatives Iowa Corp.; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

Table C.2. Estimated impact of the CHIIC intervention on selected Medicare FFS expenditures (dollars PBPM) and utilization measures, for beneficiaries enrolled within versus after 9 months of program start date

	Beneficiaries enrolled within 9 months of program start date					Beneficiaries enrolled after 9 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percent-age impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percent-age impact[a]	*p*-value
Total expenditures ($ PBPM)[b]										
Baseline year	1,249	1,068				926	959			
Year 1	1,067	1,213	-327* (193)	-24%	0.09	1,156	1,134	56 (73)	5.2%	0.44
Year 2	1,186	1,290	-285 (216)	-22%	0.19	1,221	1,234	20 (80)	1.9%	0.80
Acute inpatient expenditures ($ PBPM)										
Baseline year	480	366				297	300			
Year 1	342	398	-169 (111)	-35%	0.13	351	357	-2.1 (44)	< 1%	0.96
Year 2	423	385	-75 (126)	-18%	0.55	367	386	-15 (42)	-4.8%	0.72
Hospital outpatient expenditures ($ PBPM)										
Baseline year	281	249				311	301			
Year 1	334	318	-16 (68)	-4.6%	0.82	373	355	8.6 (16)	2.4%	0.60
Year 2	312	376	-97 (66)	-25%	0.15	403	403	-10 (22)	-2.6%	0.65
Professional Part B expenditures ($ PBPM)										
Baseline year	203	205				151	153			
Year 1	214	202	14 (34)	7.4%	0.68	183	173	12 (10)	7.3%	0.24
Year 2	221	212	11 (29)	6.2%	0.70	198	187	13 (13)	7.5%	0.31
Primary care visits in all settings, per 1,000 beneficiaries										
Baseline year	9,684	8,254				7,964	7,959			
Year 1	9,533	7,588	515 (624)	5.8%	0.41	8,187	7,503	679*** (214)	9.2%	< 0.01
Year 2	9,837	8,473	-66 (704)	< 1%	0.93	8,445	8,101	339 (266)	4.3%	0.20
Specialist visits in all settings, per 1,000 beneficiaries										
Baseline year	7,952	7,120				6,899	6,733			
Year 1	8,193	7,326	36 (813)	< 1%	0.97	7,365	7,536	-337 (262)	-4.4%	0.20
Year 2	8,030	8,211	-1,012 (1,092)	-12%	0.35	7,184	7,800	-782*** (296)	-10%	< 0.01

Table C.2 (*continued*)

	Beneficiaries enrolled within 9 months of program start date					Beneficiaries enrolled after 9 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percent-age impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percent-age impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries[b]										
Baseline year	537	471				327	342			
Year 1	381	416	-100 (99)	-21%	0.31	364	373	6.4 (29)	1.9%	0.82
Year 2	417	467	-116 (94)	-24%	0.22	368	398	-14 (33)	-4.3%	0.67
Percentage with a 30-day readmission, among all discharges										
Baseline year	7.4	13				11	13			
Year 1	11	15	2.0 (6.2)	23%	0.75	17	21	2.5 (2.5)	18%	0.31
Year 2	13	14	5.6 (7.2)	74%	0.44	12	18	0.57 (2.7)	4.9%	0.83
ED or observation visits, per 1,000 beneficiaries										
Baseline year	820	634				585	615			
Year 1	791	759	-154 (137)	-16%	0.26	648	645	33 (41)	5.4%	0.42
Year 2	927	842	-101 (161)	-10%	0.53	618	694	-46 (49)	-7.2%	0.35
Percentage of beneficiaries with any hospital admission in a time period[c]										
Baseline year	37	30				23	23			
Year 1	26	20	5.5 (5.4)	27%	0.31	22	22	-0.07 (4.1)	< 1%	0.99
Year 2	23	24	-1.1 (5.5)	-4.5%	0.85	21	22	-0.46 (4.2)	-2.1%	0.91
Percentage of beneficiaries with any ED or observation visits in a time period[c]										
Baseline year	45	37				35	36			
Year 1	39	39	0.28 (6.8)	< 1%	0.97	33	32	0.90 (5.7)	2.8%	0.88
Year 2	36	40	-4.5 (6.9)	-11%	0.52	34	35	-0.68 (5.8)	-1.9%	0.91
Sample sizes										
Number of beneficiaries	179	671				1,745	6,889			

Source: Mathematica's analysis of information from the awardee's program encounter database from September 2014 through August 2017 and Medicare claims and enrollment data from September 2013 through February 2018, as of March 13, 2019.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the

Table C.2 (*continued*)

binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiary characteristics and whether the beneficiary had any hospital stay or ED visit at baseline. The intervention years are beneficiary-specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

[c] Baseline values are the proportion of beneficiaries with more than 2 ED or observation visits in the baseline year.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CHIIC = Catholic Health Initiatives Iowa Corp.; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for Catholic Health Initiatives Iowa Corp. (CHIIC) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, relying on results from awardees with background characteristics similar to CHIIC. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on three core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for CHIIC led to a Bayesian estimate of the program's impact on total Medicare expenditures of less than -1 percent (an estimated reduction of $3 per beneficiary per month) in the first year and an impact of less than 1 percent (an estimated increase of $7 per beneficiary per month) in the second year.

Table D.1. Comparison of frequentist and Bayesian impact estimates for CHIIC in the first two years after enrollment

		Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Follow-up period	Frequentist	Bayesian	Frequentist	Prior	Bayesian
Total expenditures ($ PBPM)	Year 1	19 (-115, 153)	-2.6 (-69, 61)	2%	2%	> -1%
	Year 2	-13 (-160, 133)	6.7 (-65, 75)	-1%	3%	< 1%
Hospital admissions	Year 1	-3.7 (-58, 51)	-0.11 (-22, 21)	-1%	2%	> -1%
	Year 2	-26 (-87, 36)	3.0 (-20, 25)	-7%	3%	< 1%
ED visits	Year 1	15 (-62, 92)	-3.1 (-42, 35)	2%	2%	> -1%
	Year 2	-50 (-143, 43)	2.1 (-40, 44)	-7%	2%	< 1%

Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2016, to March 31, 2017. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

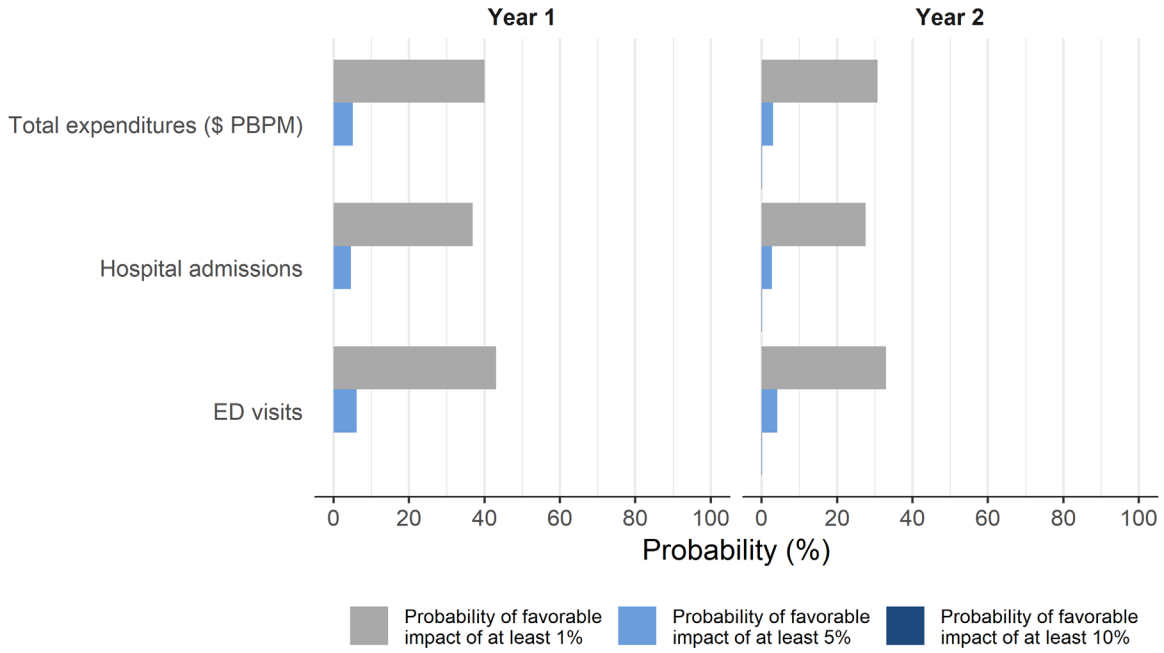
ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results relied on a small sample and are therefore imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these

differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that CHIIC achieved favorable impacts during each of the first two years on three core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the CHIIC program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from awardee's randomization file and Medicaid claims and encounter data from April 1, 2016, to March 31, 2017. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a modest probability—in the range of 40 percent—that CHIIC had a favorable impact of 1 percent or more on total Medicare expenditures, hospital admissions, and emergency department visits. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the CHIIC program did not have a meaningful impact on total expenditures or service utilization.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Children's Home Society of Florida

September 2020

Brian Bruen, Genna Cohen, Margaret Coit, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

CHILDREN'S HOME SOCIETY OF FLORIDA

Children's Home Society of Florida received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to support patient navigation and direct health care services in Pine Hills, Florida, a community near Orlando. The awardee expanded an existing patient navigation program to offer these services to the entire Pine Hills Community at the Evans Community School and a community-based office. The expanded program was known as the Pine Hills Wellness Program. To facilitate access to care, the awardee also supported direct provision of primary care and behavioral and dental services at the school and a clinic located on the school campus. The program sought to improve access to health care for residents of Pine Hills, a target population that included mostly racial minorities, many low-income individuals, many Medicaid enrollees, and a larger share of children than in the U.S. population as a whole. The program launched in October 2014, one month after receiving the HCIA R2 award. The intervention period funded by HCIA R2 ended on September 30, 2017. Table 1 summarizes key characteristics of the program.

Important issues for understanding the evaluation

- The Children's Home Society aimed to improve access to health care for residents of Pine Hills, Florida, through patient navigation and providing medical, dental, and behavioral health services in community-based settings.
- This report relies on data submitted by the Children's Home Society through the first half of the award period, before the program transitioned to new clinical partners at the Evans Health and Wellness Center in mid-2016.
- An impact analysis for this program was not feasible because there was not an administrative data source that could be used to construct a comparison group, and the data that was available did not have key information on costs or ED use. As a result, the descriptive results presented in this report should not be interpreted as measuring the causal impact of the program.

The awardee hypothesized that providing both patient navigation and direct health care services would reduce health care costs, increase the use of appropriate health care services, and improve health outcomes. The goals of the program were to (1) reduce health care costs; (2) reduce the use of emergency departments (EDs) and crisis care; (3) expand access to preventive, primary, dental, and behavioral health care services; and (4) reduce care gaps for adolescents through increased visits to primary care practitioners and improved access to preventive care.

Several factors prevented a rigorous impact analysis of this program. First, lack of Medicaid claims and enrollment data made it difficult to identify the treatment group and a credible comparison group in administrative records. Second, data provided by the program's clinical partner did not include information on health care costs or inpatient and ED service use (key outcomes of the study). Also, these data did not cover the full award period nor contain data that could be used to construct a comparison group. Finally, few individuals had enough exposure to intervention services to have a reasonable chance of achieving program impacts. As a result, the descriptive results presented in this report should not be interpreted as measuring the causal impact of the program.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The Children's Home Society of Florida received HCIA R2 funding to implement the Evans Health and Wellness Center and the Pine Hills Wellness Project, a multipronged program intended to improve access to health care for individuals living in Pine Hills, Florida.
Major innovation	The Community School model is a nationally recognized innovation to integrate community partners and provide comprehensive services that promote health and well-being. The awardee developed a program within the Community School model to expand access to health-related services to the students and others affiliated with the school and to the community at large.
Program components	• Patient navigation • Integrated medical, dental, and behavioral health care services
Target population	All residents of Pine Hills, Florida, particularly students at the Evans Community School and their families and caretakers
Total enrollment	The awardee enrolled a total of 6,017 participants (401 percent of its original goal). There was no formal enrollment process; all patients who received care at the school, the on-campus clinic, or the community office were considered to be enrolled in the program.
Level of engagement	Anyone who used services at least once was considered enrolled in the program. Data for a limited number of enrollees who used services from October 2014 to March 2016 indicated that levels of service use varied, with some individuals using services once and others engaging with providers five or more times during the 18-month period.
Theory of change or theory of action	The Children's Home Society hypothesized that implementing a program that incorporated patient navigation and direct health care services in community-based settings would lead to lower cost of care, better use of appropriate services, and enhanced patient outcomes.
Award amount	$2,078,295
Effective launch date	The program began operating in October 2014.
Program settings	• The Evans Community School • An on-campus health clinic • A community-based program office
Market area	Urban
Market location	Pine Hills, Florida
Target outcomes	• Decrease in ED visits per beneficiary • Decrease in percentage of participants with asthma who have one or more ED visits • Increase in percentage of Medicaid/CHIP population receiving timely health care • Improve participants' experience with care • Decrease in percentage of female participants younger than 18 who are pregnant • Decrease in percentage of students at Evans Community School who report risky health behaviors • Decrease in total cost of care
Payment model	A PBPM care coordination fee to be paid by Medicaid managed care organizations after end of award
Sustainability plans	At the end of the award, Children's Home Society planned to sustain some aspects of the program and to continue operating a similar program in a neighboring community. The awardee was not successful in getting payers to financially support the Evans Health Center and Pine Hills Wellness Project.

CHIP = Children's Health Insurance Program; ED = emergency department; HCIA R2 = Round 2 of the Health Care Innovation Awards; PBPM = per beneficiary per month.

PROGRAM DESIGN AND ADAPTATION

The Children's Home Society's program included two main components: (1) using patient navigators to help families, students, and other community members access needed health and social services; and (2) directly providing medical, dental, and behavioral health care services.[1] The awardee did not significantly change its service delivery model during the three-year agreement.

Patient navigation

Patient navigators helped their clients to lead healthier lives by engaging with clients and identifying their social and health care needs while supporting their efforts to pursue and adhere to appropriate care. Navigators connected students, families, and community members to a variety of health and social services, including housing and employment supports, child care, food pantries, and services that linked individuals to health insurance.

Providing medical, dental, and behavioral health care

The Children's Home Society provided behavioral health services directly to students at the school. A clinical partner offered primary care and behavioral and dental services to students, teachers, families and other community members at the Evans Health and Wellness Center, a freestanding building located at the school.[2] The provision of behavioral, medical, and dental services to participants went largely as planned. However, the awardee switched to a new clinical partner in the third program year, which resulted in the temporarily cessation of medical services and the elimination of dental services.

Student ambassadors

In addition to the two main components, the awardee also supported an optional, weekly, after-school education program for students called the Student Health Ambassadors program. This program provided information on a variety of health topics. It also sponsored field trips to health-related organizations—for example, to a hospital or medical school—to promote greater engagement by students in their own health and the health of their families and community, as well as to encourage potential interest in health-related careers.

[1] The Third Annual Evaluation provides additional details on the design and implementation of the Children's Home Society. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

[2] The HCIA R2 funding was one of several funding streams that supported activities at the school and the Evans Health and Wellness Center. The awardee leveraged funding beyond what it received from HCIA R2 to support the activities offered in the school. Each funding stream was necessary to achieve the awardee's goals and it was not possible to assess the differential impact of each funding stream.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The Children's Home Society experienced several implementation challenges: (1) an unexpected change in clinical partners, (2) challenges related to operating a health program on a school campus, (3) lack of trust in government organizations from within the community, (4) the limited resources of community residents, and (5) safety issues in the surrounding neighborhood. The most disruptive challenge was the transition in clinical partners at the on-campus clinic. The original clinical partner struggled to adapt its service provision model to the needs of the student and community populations in a financially sustainable way. The first clinical partner ramped down enrollment of new patients near the end of the second program year (August 2016) and stopped providing services mid-way through the third program year (December 2016). A new clinical partner had not yet become fully operational at the on-campus clinic by the end of the cooperative agreement. In addition, the co-location of the clinic on school grounds facilitated student access, but it also limited access among community residents. Florida law prohibited students and community residents from visiting school-based clinics during the same hours. Despite these initial challenges, the program ultimately succeeded in establishing clinic times dedicated for students and focused on increasing buy-in and participation from teachers, which enhanced students' abilities to access services.

The awardee faced challenges in engaging community members as well. The target community included many low-income households. Residents often lacked the resources (such as transportation, funds for bus fares, and flexibility to leave work) to attend preventive or follow-up health appointments at the on-campus clinic. A large immigrant population in Pine Hills with limited trust of government and community service institutions hindered efforts by patient navigators to recruit them into the program. In addition, there was no direct route from the bus depot to the on-campus clinic, requiring community residents without other transportation options to walk through a dangerous neighborhood to access services. Adding to these challenges, substantial turnover among health navigators in the second program year further hindered recruitment of community residents.

To resolve these community engagement issues, the awardee hired new navigators with more experience working with the residents, including some navigators who

Implications of program implementation for achieving program goals

- The Children's Home Society used its HCIA R2 funds to expand existing initiatives supported through multiple funding streams. This made it difficult to define the intervention and inhibited identifying comparison groups.
- The Children's Home Society did not differentiate between students and their families versus other community residents in available data, so it was not possible to separately track service use for these different populations.
- Due to the transition in clinical partners and lack of availability of timely data, the descriptive analysis does not include data after March 2016.

lived in the community. By the third program year, staff reported that the program was adequately staffed, and staff turnover was no longer a problem. The awardee also began to use more efficient recruitment methods, leveraging partnerships with businesses and institutions already trusted by the community residents.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Enrolling and engaging participants

Despite the implementation challenges it faced, the awardee enrolled 6,017 participants by the end of the three-year cooperative agreement, more than four times the original enrollment projection. The awardee considered any person who received patient navigation or obtained health care services at the Evans Health and Wellness Center an enrollee. The awardee was mostly successful in engaging students, especially through the Student Health Ambassadors program. But staff reported barriers connecting with and maintaining trust among community members limited residents' engagement throughout the award period. Nonetheless, staff believed that the program made a difference in meeting the needs of the community and had a positive impact on participants' health goals.

Characteristics of program participants and service use

The descriptive analysis in this report is limited to two samples that represent a subset of the 6,017 enrollees. One of these samples contained 99 participants and the other contained 1,750 participants, which meant that data was available for at most 1,849 participants. [3] No characteristics were available for the remaining participants (about 70% of enrollees).

First, the Children's Home Society provided data for 99 participants who received either behavioral health services or patient navigation services from its own staff from October 2014 to March 2016. The awardee's behavioral health providers were embedded in the school, and nearly all recipients in this data file were younger than 18 at the time of service. More than half (56) were female and nearly all were Black or African American, Hispanic, or another non-white, non-Hispanic races and ethnicities. Attention deficit disorder with hyperactivity was the most common diagnosis on the behavioral health claims for these participants. Other common diagnoses included adjustment disorders with varying symptoms (for example, disturbance of emotions and conduct and mixed anxiety and depressed mood) and adolescent-onset conduct disorder. These conditions were consistent with the targeting of services to a student population.

The 99 participants in the Children's Home Society data file obtained a mix of behavioral health and patient navigation services. Therapeutic behavioral care was the most commonly used service. Mental health assessments, mental health service plans, and comprehensive medication

[3] It was not possible to link individuals between the two samples, so a small number of people may be counted twice across these two samples.

services were also common behavioral health procedures. Nearly half of this sample of participants had five or more visits during the 18-month observation period, suggesting a fairly high level of engagement with the behavioral health and patient navigation providers. The number of procedural codes during a visit also suggests that the small subset of youth for whom data were available received a high volume of patient navigation services. The procedural codes did not identify the types of services provided.

The second data set used for the descriptive analysis contained medical, dental, and behavioral health claims for 1,750 participants who received services at the on-campus clinic from October 2014 through March 2016. These data provide a more complete picture of the beneficiaries served by this program, but they include only individuals who obtained care at the on-campus clinic and do not cover the entire three-year award period. Nearly 40 percent of these participants were children (younger than 18) and the overwhelming majority were Black or Hispanic. Although there were roughly equal numbers of male and female children using the on-campus clinic, 70 percent of the adults who used the clinic were female.

Main findings from descriptive analysis

- The diagnoses, procedures, and patterns of care were consistent with the limited size of the Evans Health and Wellness Center and its focus on primary care, preventive care, and general and nonsurgical dental care for a predominately low-income population.
- These services also illustrate a need within the community for basic health care (for example, office visits, immunizations, and dental checkups) and prevention.
- However, the relatively low number of repeat users could indicate the transportation or other resource challenges that many residents faced or the lack of ongoing engagement with community members.

Among the claims for the 1,750 participants seen at the on-campus clinic, common activities and procedures included standard patient office visits, infant or child health checks, asthma-related care, immunizations, routine venipuncture (for example, drawing blood for tests), dental examinations and fillings for caries (cavities), and comprehensive preventive medicine evaluation and management for new patients. Nearly half of the participants (47 percent) had just one visit to the Evans Health and Wellness Center from October 2014 to March 2016, whereas 13 percent had five or more visits. The records also show that many patients had insurance, including Medicaid, Medicare, and private coverage. However, about one-third of all claims were expected to be paid on a sliding fee scale based on income or noted that no payment was expected, which indicated that these participants did not have health insurance.

Challenges of measuring program impacts

As noted earlier in this report, a rigorous impact analysis of this program was not possible due to the inability to use Medicaid claims data or another administrative data source to identify the treatment group and construct a credible comparison group. The administrative data that was

available was only available for the treatment group and was missing key information on outcomes. Finally, few individuals had enough exposure to intervention services to have a reasonable chance of achieving program impacts. Therefore, the descriptive results in this report cannot be used to draw inferences about the program's impact on outcomes.

CONCLUSION

Due to data limitations, it was not possible to assess the impact of the program on health care service use and costs. However, a review of qualitative and program data supplied by the awardee indicated that the Children's Home Society was partly successful in implementing its community-based patient navigation intervention and providing direct health care services to the target population of low-income students and residents. Records show that students received patient navigation services, mental health assessments, service planning, and behavioral health treatment services at the school, with many remaining engaged in services for extended periods. Records also show that the medical and dental services provided at the on-campus clinic were consistent with the needs of a low-income community. Program staff also reported that they believed the community-based program made a difference in meeting the needs of the residents and had a positive impact on participants' health goals.

PROGRAM SUSTAINABILITY

After its award ended in August 2017, the Children's Home Society eliminated some aspects of is program while sustaining others by transferring responsibilities to its partners. Closing the Pine Hills Wellness Project office was a significant cut, as it discontinued community-based patient navigation and outreach services. The Children's Home Society also ended sponsorship of the Student Health Ambassadors program. Although participants wanted to continue the student-peer health advocacy and engagement activities, there was uncertainty about whether the public school system, a grant of some kind, or other means could sustain the program. The awardee's second clinical partner, Orange Blossom Family Health, took over responsibilities for providing the full array of medical, dental, and behavioral health services to the Evans Community School students and staff and members of the community, primarily through insurance billing and other health

Children's Home Society's proposed payment model

The Children's Home Society envisioned a payment model in which Medicaid managed care organizations (MCOs) pay a per beneficiary per month (PBPM) fee for the awardee and its partners to provide primary and preventive medical, dental, and outpatient behavioral health care, as well as community health and wellness promotion activities. Under the proposed payment model, the awardee's clinical partner would receive the PBPM fees and provide the covered primary and preventive care services. The awardee expected to determine the amount of the PBPM fee when negotiating with MCOs, but did not enter such negotiations before the award ended.

center funding streams. Orange Blossom's staff also assumed responsibility for most of the school-based health navigation services, with the awardee maintaining some of those services using funding from a legislative appropriation from the state.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: City of Mesa Fire and Medical Department

September 2020

John Schurrer, KeriAnn Wells, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

CITY OF MESA FIRE AND MEDICAL DEPARTMENT

The City of Mesa received funding under Round 2 of the Health Care Innovation Awards (HCIA R2) to develop the Community Care Response Initiative (CCRI). The CCRI aimed to address several problems associated with using the 911 emergency response system for non-emergency care cases. The CCRI specifically sought to address (1) the high cost of transporting people with low-acuity conditions by ambulance and treating them in hospital emergency departments (EDs), (2) diverting professionals and resources needed to respond to true emergencies, and (3) overcrowding EDs with patients who do not need emergency care.

The CCRI addressed these problems by dispatching community medicine (CM) units to provide non-emergency services as appropriate to 911 callers at their homes or in the community. The awardee introduced two types of CM units—CM medical units and CM behavioral units—to respond to low-acuity 911 calls, choosing the appropriate unit based on the callers' needs. To complement the 911 response component of the CCRI, which focused on low-acuity cases, CCRI added a care transitions (CTs) component that provided home-based services to higher-acuity participants with targeted conditions after discharge from a hospital.

Important issues for understanding the evaluation

- The impact evaluation focused on the program component that aimed to divert low-acuity 911 callers from the ED by dispatching a mobile health unit instead of an ambulance in response to the 911 call.
- The model was an expansion of a pilot program that the Mesa area had previously implemented.
- The main impact analysis relied on 2,872 Medicaid beneficiaries who placed a 911 call to the Mesa dispatch center, received a mobile CM unit visit or had an ambulance transport, and met the inclusion criteria for the impact analysis.
- An estimation of impacts on 1,750 Medicare fee-for-service (FFS) beneficiaries who met the same evaluation eligibility criteria supplements the Medicaid analysis.

The awardee hypothesized that the CCRI could (1) reduce low-risk patients' ED and ambulance use by 40 percent, (2) reduce high-risk patients' hospital readmissions, and (3) reduce total health care spending by $41 million. The awardee expected cost savings to accrue entirely from the 911 response component; the CCRI did not develop cost savings goals for the CT component. The program launched in December 2014 and ran through February 2018, after a six-month no-cost extension. Table 1 summarizes the program's key characteristics.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The Mesa Fire and Medical Department redesigned its 911 emergency response by dispatching CM units to low-acuity callers and treating them on site in the community—often at home—instead of transporting them to the ED. CM units also provided CT services to participants with selected chronic conditions who were recently discharged from the hospital.
Major innovation	The CCRI added non-emergency clinical response teams into 911 dispatch protocols.
Program components	• CM component: Using medical units staffed with an APP and a paramedic or behavioral units staffed with a licensed behavioral health clinician and a paramedic to provide direct care to 911 callers whose conditions did not warrant ED visits • CT component: Conducting home visits and care coordination for high-acuity participants within 72 hours of hospital discharge
Target population	• CM component: Low-acuity 911 callers • CT component: Patients with CHF, COPD, MI, sepsis, and pneumonia who were recently discharged and identified as being at high risk for readmission
Participating providers	The City of Mesa Fire and Medical Department in partnership with Mountain Vista Medical Center and Crisis Preparation and Recovery, which respectively employ the APPs and behavioral health clinicians who, with paramedics, provided community medicine services to low-risk 911 callers
Total enrollment	The CCRI reported it served 12,818 patients from December 2014 through February 2018, 47 percent of its original three-year goal.[a]
Theory of change or theory of action	The awardee hypothesized that (1) using CM units to treat low-acuity participants on site in the community would decrease inappropriate use of both the ED and ambulances to transport participants to the ED, and (2) using CM medical units to provide higher-acuity participants with CT services in their homes after hospital discharge would result in fewer readmissions to the hospital. This approach would help reduce ED overcrowding, focus emergency services on priority patients, and reduce hospital readmissions, thus reducing costs and improving the quality of care.
Award amount	$12,779,725
Effective launch date	December 1, 2014
Program settings	• CM component: Medical and behavioral units dispatched from Mesa Fire and Medical Department stations to community settings and participants' homes • CT component: Medical units dispatched from Dignity Health to participants' homes
Market area	Urban and suburban
Target outcomes	• Reduce low-acuity patients' ED visits and ambulance use by 40 percent in three years • Reduce ambulance use and ED visits to save $41 million in three years • Reduce high-acuity patients' hospital readmissions in three years (high-acuity patients include those diagnosed with CHF, COPD, MI, pneumonia, and sepsis)
Payment model	New FFS payment
Sustainability plans	The awardee stopped operating its program in February 2018 due to lack of funding. A few months before the award ended, the awardee discontinued the CT component due to low enrollment and an interest in focusing on the core intervention.

[a] The awardee reported the number of beneficiaries served by the program. This number does not link to those that to the evaluation identified using the finder file provided by the program.

APP = advanced practice provider; CHF = congestive heart failure; CM = community medicine; COPD = chronic obstructive pulmonary disease; CT = care transition; ED = emergency department; FFS = fee-for-service; MI = myocardial infarction.

The impact analysis presented in this report includes only the 2,872 Medicaid and 1,750 Medicare FFS beneficiaries who called 911 from December 2014 through February 2018, received either a CM medical or behavioral health unit visit or were transported to the ED via ambulance, and met the claims and eligibility inclusion criteria for the evaluation. This represents 41 and 33 percent of the Medicaid and Medicare beneficiaries, respectively, from the awardee's finder file. An impact evaluation of the CT component aimed at reducing readmissions among high-acuity patients was not possible due to low enrollment. The comparison group included 11,291 Medicaid comparison beneficiaries and 6,014 Medicare comparison beneficiaries who had similar baseline demographics, health status, and service use; lived outside the area served by the CCRI; and who had an ambulance transport. Table 2 summarizes the key features of the evaluation. Appendix A, Tables A.1 and A.2, describe the identification of the analytic sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group.
Intervention group for evaluation	The study sample included 2,872 Medicaid beneficiaries who placed a 911 call to the Mesa dispatch center from December 2014 through February 2018, received a mobile CM unit visit or had an ambulance transport, and met the inclusion criteria for the impact analysis. Of the 2,872 Medicaid beneficiaries, about 75 percent had a CM visit. The study sample also included 1,750 Medicare FFS beneficiaries who met the same eligibility criteria (of whom 44 percent had a CM visit).
Comparison group	The comparison group included matched sets of 11,291 Medicaid beneficiaries and 6,014 Medicare FFS beneficiaries with similar demographic, health, expenditure, and service use characteristics as the treatment beneficiaries. The comparison group beneficiaries met the same eligibility criteria as the intervention group, lived in geographic regions similar to the Mesa 911 service area, and received an ambulance transport to an ED.
Limitations	If participants differed from eligible nonparticipants in ways not captured in Medicare administrative files and claims, the impact estimates might be biased. A specific concern is the inability to classify the reason for dispatch in the comparison group and then match on this, which could introduce concerns about unobservable characteristics that might affect future service use and expenditures.

CM = community medicine; ED = emergency department; FFS = fee for service.

PROGRAM DESIGN AND ADAPTATION

As part of the CM component, the Mesa Fire and Medical Department hired two triage nurses to help 911 dispatchers identify low-acuity callers and decide when to deploy CM units. The 911 dispatchers followed a decision tree to determine whether to dispatch a medical or behavioral health unit, or an ambulance. Dispatchers also could transfer callers to the triage nurse, who asked additional questions and dispatched the appropriate response, or added a unit to the response already dispatched. (In spring 2017, the awardee discontinued the nurse triage function due to low demand for its services.) Staff on CM units also communicated with triage nurses and callers en route and listened to calls coming over the radio. If the CM unit was available and

responders thought it might benefit a caller, they could go to the caller's location with emergency responders and assess the situation on the spot.

CM medical units used ambulances staffed with a paramedic and an advanced practice provider (APP), who provided on-site services like those in an urgent care setting, such as suturing wounds and administering antibiotics. These are services beyond what a paramedic can deliver on site. The APPs were typically nurse practitioners but a handful of physician assistants also were available. Low-acuity medical issues commonly included self-limiting illnesses (such as flu), minor injuries requiring sutures, managing diabetes or nose bleeds, or conditions requiring antibiotics. At program launch, the awardee operated three CM medical units.

CM behavioral units used sport utility vehicles staffed with a paramedic and a licensed behavioral health clinician who provided non-emergency behavioral health or crisis intervention services for conditions including anxiety, depression, substance abuse, and suicidal ideation. CM behavioral unit visits often lasted a few hours as behavioral health clinicians assessed participants and worked to find the most appropriate treatment venue. Many behavioral health calls resulted in transporting patients directly to local inpatient behavioral health facilities, bypassing the ED. At program launch, the awardee deployed three CM behavioral units.

Implications of program implementation for detecting impacts

- The goal of the program was to divert low-acuity patients from an ED, so any impacts are likely to concentrate on ED visits and appear shortly after enrollment.
- The documentation challenges that resulted from the systems used and changes to the systems over time affected the ability to identify participants in Medicare and Medicaid enrollment and claims files.

Throughout the cooperative agreement, the awardee reviewed internal monitoring and tracking data on the CCRI to continuously refine its dispatch and treatment protocols. For example, dispatchers stopped sending CM medical units to patients with abdominal pain because they could not divert most of those patients from the ED. Importantly, 911 dispatchers typically did not dispatch CM units to older callers because they considered age as a factor in determining acuity. In April 2017, in response to attrition among APPs and to concentrate services during peak hours, the awardee changed CM medical units' hours of operation. CM units transitioned from providing services 24 hours a day, seven days a week, to operating on a more limited schedule.[1]

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

This program was an expansion on an existing pilot program and the awardee quickly began start-up activities, which included developing operational plans, securing CM vehicles, and hiring APPs and behavioral health clinicians. This enabled the awardee to launch the CM component in the second quarter after entering the agreement. CCRI successfully recruited 11 APPs and 4 behavioral health clinicians when it launched the CM component, achieving 88 percent of its staffing target by the end of the first year of the cooperative agreement. In the second year of the cooperative agreement, the awardee hired an additional APP and a behavioral health clinician.

Throughout the agreement, staff and clinicians repeatedly expressed their support for the CM component as a better model for providing care. Buy-in among paramedics and clinicians was a major factor in the program's implementation success, enabling them to promote the program to participants and community partners, such as the police. Staff retention was strong until the final year of the cooperative agreement, when some APPs left because they did not know whether their employment would continue after the end of the agreement. As APPs left the program, the awardee reduced the number of CM medical teams, which meant that captain paramedics who had been paired with APPs returned to traditional emergency response services.

Despite the staffing successes, some fire department station chiefs expressed concerns about having clinicians in their station who did not report to them. Captain paramedics also noted that all APPs did not have the same skill set in the field. In response, the awardee began having clinicians report to battalion chiefs for certain operational issues in the second year of the award. In the third year of the award, the awardee offered APPs a two-day training to develop uniform competency on core services. The awardee also struggled to integrate the triage nurses into the CM unit dispatch protocol and ultimately discontinued the nurse triage function in spring 2017.

Documentation also proved challenging for the awardee. At first, the awardee documented 911 response encounters in an encounter-based emergency medical services (EMS) tracking system. However, this system was not conducive to tracking and billing participants. In May 2016, the awardee implemented an electronic medical record (EMR) for APPs, while paramedics continued to document encounters in the existing EMS system. However, the awardee continued to face challenges with the new EMR and found it was not conducive to documenting care delivered by the CM units. In December 2016, the awardee reverted to its original system, with enhancements to accommodate billing, enabling APPs and paramedics to use the same system.

The awardee also indicated that the zero-dollar copayment for Medicaid ED services in Arizona sometimes affected a beneficiary's decision to use ambulance transport or seek additional ambulatory services. For example, the state Medicaid agency initially paid for cab rides to the ED, but later reversed course and reimbursed only for an ambulance transport. Low-acuity beneficiaries without transportation options would therefore opt for transport in the ambulance.

The awardee also reported that when the CM unit wanted to administer an antibiotic and suggested the patient see a primary care provider, some patients preferred to go to the ED because it did not charge copayments, but physician visits required copayments.

ESTIMATING PROGRAM IMPACTS

Study sample

The awardee's finder file included 113,806 unique 911 callers from December 2014 through February 2018, of whom 7,015 were linked to Medicaid enrollment data and 5,266 linked to Medicare enrollment data. The documentation challenges described previously affected the ability to identify beneficiaries in enrollment and claims data.

After applying the claims- and enrollment-based sample selection criteria, the analytic sample included 2,872 Medicaid and 1,750 Medicare beneficiaries identified from 911 callers to the Mesa dispatch center. The study dropped most of the excluded Medicaid beneficiaries because they lacked Medicaid enrollment in the 90 days before their 911 calls (2,518 beneficiaries) or on the day of enrollment (763 beneficiaries). The study dropped another 593 beneficiaries because they did not receive a CM visit as recorded in the data provided by the CCRI or an ambulance transport in Medicaid claims data. And, it dropped 261 beneficiaries for other eligibility reasons, such as they had restricted benefits or were dually eligible for Medicare and Medicaid—the Medicare analytic sample included dually eligible beneficiaries. The study dropped 8 beneficiaries during the matching process because it could not identify good comparison beneficiaries.

Among the 5,266 911 callers who could be linked to Medicare enrollment data, 2,453 beneficiaries were dropped because they were enrolled in a Medicare advantage plan. Another 668 beneficiaries did not have Medicare Parts A and B, Medicare as the primary payer or 90 days of FFS coverage during the baseline period. And, the study dropped 372 beneficiaries because they did not receive a CM visit as recorded in the data provided by the CCRI or an ambulance transport in Medicare claims data. The study dropped 23 beneficiaries during the matching process because it could not identify good comparison beneficiaries Appendix A provides more details.

Because it was not possible to replicate the CCRI's eligibility criteria using claims to select a low-acuity comparison group, the study used an intent-to-treat (ITT) analysis, which measures the impact of the CCRI on Medicaid and Medicare beneficiaries who called 911 and received either a CM visit or an ambulance transport relative to Medicaid and Medicare beneficiaries who had an ambulance transport. Because the CCRI's intent was to dispatch the CM units for low-acuity cases, it does not appear that the dispatchers were successful at doing so among Medicare beneficiaries. Of the 2,872 Medicaid beneficiaries included in the analysis, about 75 percent received a CM visit; among the 1,750 Medicare FFS beneficiaries for whom the study estimated impacts, 44 percent received a CM visit.

Estimating impacts for all 911 callers transported to an ED reduces the risk of biased impact estimates that might be caused by the inability to replicate the dispatch criteria for the selection of a comparison group when there are systematic differences in baseline health status and service use between those receiving a CM visit and those transported by ambulance to an ED. Appendix B presents evidence of differences by comparing the characteristics of the 911 callers who received a CM visit and those callers who did not receive one. Medicaid beneficiaries who received a CM visit had, on average, lower risk scores and lower levels of baseline service use and disease burden compared to those transported by an ambulance to the ED. Conversely, Medicare beneficiaries who received a CM visit had, on average, higher risk scores, greater baseline expenditures and service use, and were more likely to be dually eligible for Medicare and Medicaid. The factors behind the differences are likely due to characteristics that cannot be fully identified in claims, such as the reason for the 911 call and social determinants of health. This evidence supports an ITT analysis based on all 911 callers rather than just those that who received a CM visit.

Matched sets of 11,291 Medicaid and 6,014 Medicare FFS beneficiaries who met the same enrollment- and claims-based eligibility criteria as the treatment group served as the comparison groups. Comparison beneficiaries also had to live outside the Mesa 911 catchment area in regions of Arizona that had characteristics similar to Mesa's catchment area and had a claim for an ambulance transport that likely originated with a 911 call during the study period. Appendix B provides the full balance results measured during the 12 months before enrollment.

Characteristics of treatment and comparison group beneficiaries

Medicaid beneficiaries

The study achieved reasonable balance on most baseline characteristics between the Medicaid treatment beneficiaries and the matched comparison group (Table 3). More than half of the treatment and comparison beneficiaries were ages 21 to 50. A number of beneficiaries lived with a disability—18 percent of the treatment beneficiaries and 21 percent of the comparison beneficiaries. Both Medicaid treatment and comparison beneficiaries had conditions that led to predicted expenditures that were more than twice the average for all Medicaid beneficiaries, as measured by the Chronic Illness and Disability Payment System (CDPS) risk score.[2] Psychiatric, pulmonary, and cardiovascular conditions were prevalent.

The treatment group had about 10 percent greater use of hospitalizations and ED visits than the comparison group during the year before enrollment. Given the disease burden and the relatively large proportion of beneficiaries living with a disability, it is unsurprising that both treatment and comparison beneficiaries were high-volume service users. The Medicaid analysis does not include expenditure data due to the high level of Medicaid managed care in Arizona.

[2] A beneficiary with a risk score of 1 indicates that the individual is expected to have average future expenditures based on age, gender, and diagnosis history. Lower scores signify lower predicted expenditures; higher scores signify higher predicted expenditures.

Table 3. Baseline characteristics of Medicaid beneficiaries in treatment and comparison groups

Measure	Treatment (N = 2,872)	Comparison (N = 11,291)
Demographics[a]		
Age, years	35	33
Male, %	47	47
Disabled, %	18	21
CDPS score[b]		
Mean	2.27	2.22
25th percentile	0.78	0.78
Median	1.71	1.71
75th percentile	2.96	2.94
Selected conditions, %		
Pulmonary condition	26	27
Diabetes	13	13
Cardiovascular condition	31	27
Psychiatric condition	46	45
Service use and expenditures during the year before enrollment		
Number of hospital admissions (per 1,000 beneficiaries per year)	743	675
Number of outpatient ED visits and observation stays (per 1,000 beneficiaries per year)	4,376	3,927
Percentage with an outpatient ED visit in 12 months before enrollment (%)	72	72
Number of ambulance transports (per 1,000 beneficiaries per year)	1,238	1,151

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018.

Notes: The study defined the baseline year as the 365 days before each beneficiary's enrollment date. It defined the enrollment date as the date of the 911 call that triggered a potential CM unit dispatch for treatment beneficiaries and an emergency ambulance transport for comparison beneficiaries. The study measured all beneficiary characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

Appendix B presents full balance results.

[a] Racial and ethnic data in Medicaid are unreliably reported and not presented here.

[b] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

CDPS = Chronic Illness and Disability Payment System; CM = community medicine; ED = emergency department.

Medicare beneficiaries

The Medicare treatment and matched comparison group were also well balanced on most baseline characteristics (Table 4). As expected, most beneficiaries were 65 and older, but about one-fifth of the treatment beneficiaries were younger than 65. And more than 20 percent of both groups were dually eligible for Medicare and Medicaid. Like the Medicaid beneficiaries,

Medicare beneficiaries had predicted expenditures that were at least twice the national average, as measured by the hierarchical condition category (HCC) risk score.[3] There was significant disease burden in both groups, with about one-quarter of beneficiaries having chronic obstructive pulmonary disease or congestive heart failure. Diabetes was also a common condition.

Medicare beneficiaries were also high-volume users of inpatient hospital and ED services. Medicare beneficiaries had higher rates of hospitalizations, but lower rates of ED visits compared to Medicaid beneficiaries. Differences in the health needs of the Medicaid and Medicare beneficiaries could in part explain these differences. The Medicare expenditures reflect the high utilization rates, with average total Medicare expenditures of $2,184 per beneficiary per month (PBPM) for the treatment group and $2,009 PBPM for the comparison group.

Table 4. Baseline characteristics of Medicare beneficiaries in treatment and comparison groups

Measure	Treatment (N = 1,750)	Comparison (N = 6,014)
Demographics		
Age group, %		
Younger than 65	20	17
65 to 74	29	29
75 to 84	29	32
85 and older	22	22
Male, %	46	47
White, %	91	91
Original reason for Medicare eligibility, %		
Old age and survivor's insurance	69	73
Disability insurance benefits[a]	29	26
Medicare–Medicaid dual status, %	22	21
HCC score[b]		
Mean	2.1	2.0
25th percentile	0.9	0.8
Median	1.5	1.5
75th percentile	2.7	2.6
Selected conditions, %		
Chronic obstructive pulmonary disease	26	26
Diabetes	32	29
Congestive heart failure	25	24
Major depressive, bipolar, and paranoid disorders	15	11

[3] The CDPS algorithm is tailored to the Medicaid population, whereas the HCC algorithm is designed for the Medicare population. More information on the CDPS algorithm is available at http://cdps.ucsd.edu/. Additional information on the HCC algorithm is available at https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.

Table 4 (*continued*)

Measure	Treatment (N = 1,750)	Comparison (N = 6,014)
Service use and expenditures during the year before enrollment		
Number of hospital admissions (per 1,000 beneficiaries per year)	788	708
Number of ED visits and observation stays (per 1,000 beneficiaries per year)	1,883	1,606
Percentage with an ED visit in 12 months before enrollment (%)	54	54
Number of ambulance transports (per 1,000 beneficiaries per year)	989	857
Total Medicare expenditures ($ PBPM)	$2,184	$2,009
Ambulance expenditures ($ PBPM)	$28	$29

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicare claims and enrollment data from December 2012 through August 2018.

Notes: The study defined the baseline year as the 365 days before each beneficiary's enrollment date. It defined the enrollment date as the date of the 911 call that triggered a potential CM unit dispatch for treatment beneficiaries and an emergency ambulance transport for comparison beneficiaries. The study measured all beneficiary characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

Appendix B presents the full balance results.

[a] Includes participants with both a disability and ESRD.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of 1 represents average expected expenditures. The study calculated HCC scores using the most recently available HCC algorithms.

CM = community medicine; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The analysis used a difference-in-differences study design to estimate impacts. This design measures program effects as the change in outcomes among beneficiaries before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables will capture differences on unobserved variables as well.

The year before each participant's enrollment date defined the baseline period. Because the intervention sought to divert low-acuity 911 callers from the ED, the awardee hypothesized that the impacts would concentrate in the period immediately after enrollment. As a result, the analysis estimated impacts quarterly, as well as over the full year after enrollment. For treatment beneficiaries, the date of first 911 call defined the enrollment date. Each comparison beneficiary received a pseudo-enrollment date, based on the date of the claim for the beneficiary's ambulance transport.

The study chose impacts on the 2,872 Medicaid beneficiaries who called 911 as the primary analysis because of the 60 percent larger Medicaid sample size and the greater percentage of 911

callers receiving a CM visit (75 percent for Medicaid versus 44 percent for Medicare). The larger sample size improved the ability to detect impacts and the higher participation rate among Medicaid participants provided a greater ability to generalize the findings to all eligible patients with Medicaid insurance. Impacts on the 1,750 Medicare beneficiaries who called 911 supplement the main analysis.

The primary outcomes were the number of ED visits and ambulance transports for Medicaid and Medicare beneficiaries and total and emergency ambulance expenditures for Medicare beneficiaries. Other outcomes included the number of hospitalizations and primary and specialty care visits. To improve the precision of the estimates for the Medicaid analysis, outlier values for utilization measures were top-coded at the 98th percentile of the treatment group. Too few beneficiaries in the study sample received a CM behavioral health visit—87 for Medicare and 463 for Medicaid—to support separate analyses by type of mobile visit. The results presented in this report represent the combined impact of mobile visits for medical and behavioral services. Appendix A provides additional detail on the statistical methods used to estimate program effects.

Finally, this was an expansion of a pilot program and, therefore, the study did not estimate impacts separately based on the time of enrollment from the launch of the CCRI. New programs might require a ramp-up period to address implementation challenges. And although there were implementation challenges, as described earlier, the awardee established the program's foundation during the pilot phase.

IMPACT RESULTS

In line with CCRI's theory of change, the intervention reduced ambulance transports among Medicaid beneficiaries by an estimated 32 percent in the first quarter and by roughly half that amount over the full year after enrollment, both of which were statistically significant (Table 5). There were estimated reductions among Medicare beneficiaries as well, but they were smaller in magnitude over both time periods and only statistically significant in the first quarter after enrollment, with an estimated 7 percent reduction in the number of transports. Despite the smaller impacts on the number of transports for Medicare beneficiaries, the CCRI reduced expenditures for ambulance services by an estimated 24 percent during the first quarter after enrollment and by about half that amount when measured over the full 12-month follow-up period, both of which were statistically significant. The successful launch and buy-in among paramedics and clinicians might help explain the success in reducing ambulance transports and expenditures.

The estimated reduction in Medicare spending for ambulance services was not enough to lower total Medicare expenditures. The low participation rate among Medicare beneficiaries included in the ITT sample likely attenuates the effect of the program on total expenditures.

The CCRI did not have a discernible impact on the number of ED visits as measured over Medicaid beneficiaries in the first quarter, but there was an estimated 11 percent increase in ED visits over the full 12-month post-enrollment period. There was no estimated impact on ED utilization among Medicare beneficiaries. Appendix C presents the full results of the impact analysis. Appendix D provides results from a Bayesian analysis.

Table 5. Estimated impact of CM mobile unit services on selected Medicaid and Medicare outcomes

	Medicaid		Medicare	
	3 months after enrollment	12 months after enrollment	3 months after enrollment	12 months after enrollment
Number of ambulance transports, per 1,000 beneficiaries[a]				
Impact	-1,821***	-366***	-426***	58
Percentage impact	-32%	-14%	-7.3%	2.3%
p-value	< 0.01	< 0.01	< 0.01	0.51
ED visits or observation stays, per 1,000 beneficiaries[a]				
Impact	-241	553***	-210	15
Percentage impact	-3.0%	11%	4.2%	< 1.0%
p-value	0.18	< 0.01	0.24	0.88
Hospital stays, per 1,000 beneficiaries[a]				
Impact	-276***	8.7	76	158***
Percentage impact	-15%	< 1.0%	2.7%	12%
p-value	< 0.01	0.81	0.48	< 0.01
Primary care visits in ambulatory setting, per 1,000 beneficiaries[a]				
Impact	-154	100	520	349
Percentage impact	-3.0%	2.3%	4.1%	3.7%
p-value	0.27	0.39	0.27	0.22
Specialist visits in ambulatory setting, per 1,000 beneficiaries[a]				
Impact	-707***	46	633	425
Percent impact	-8.5%	< 1.0%	4.1%	3.3%
p-value	< 0.01	0.78	0.10	0.14
Ambulance expenditures ($ PBPM)				
Impact			-49***	-11***
Percentage impact			-24%	-13%
p-value			< 0.01	< 0.01
Total expenditures ($ PBPM)				
Impact			-93	174
Percentage impact			-1.3%	4.8%
p-value			0.74	0.22
Sample size				
Treatment	2,872	2,872	1,750	1,750
Comparison	11,291	11,291	6,014	6,014

Table 5 (*continued*)

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid and Medicaid claims and enrollment data from December 2012 through August 2018.

Note: Impact estimates are the regression-adjusted difference between change in the follow-up and baseline outcomes for treatment group members and the change in the follow-up and baseline outcomes for comparison group members—or the difference-in-differences estimate. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the predicted treatment group mean in the post period minus the impact estimate). Appendix C presents full impact estimates. Appendix D provides results from the Bayesian analysis. The baseline utilization rates for the Medicaid analysis do not align with those presented in Table 3 because impacts are estimated for top-coded outcomes, whereas the rates in Table 3 are not top-coded. The baseline utilization rates in Table 3 align with the rates in Appendix C, Table C.3 that present estimated impacts using non-top-coded outcomes.

[a] Outcomes top-coded at the 98th percentile for Medicaid beneficiaries. The threshold is determined among treatment beneficiaries only by pooling observations across the 12-month baseline period and the quarterly follow-up periods.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CM = community medicine; ED = emergency department; PBPM = per beneficiary per month.

Despite the success in reducing the number of ambulance transports, the difficulty in changing the behavior of frequent ED users was a factor in the persistence of high ED use after the first quarter among Medicaid beneficiaries. The awardee reported that frequent users were difficult to divert because they often preferred to go to the ED regardless of acuity. The awardee worked with beneficiaries to help them understand the potential value of a CM unit visit, but was not always successful in diverting them from the ED. A few respondents to the patient survey reflected this view by reporting that they did not like the suggestion that they did not need to go to the ED.

The zero-dollar copay for Medicaid ED services is another potential reason the initiative did not sustain estimated reductions in ED visits for the Medicaid population. As the awardee noted, beneficiaries would sometimes opt for transport to an ED because services furnished in an ambulatory setting would require a copay or transportation.

Main findings from impact evaluation

- The CCRI reduced the estimated number of Medicaid ambulance transports over both a 3- and a 12-month follow-up period.
- Medicare ambulance transports declined in the short run only, but Medicare ambulance expenditures declined over both a 3- and a 12-month follow-up period.
- There was no short-term effect on the estimated number of ED visits among Medicaid or Medicare beneficiaries, but there was an estimated increase among Medicaid beneficiaries over a 12-month follow-up year.
- The number of hospitalizations and the number of specialist visits among Medicaid recipients declined within the first three months of their 911 call. For Medicare beneficiaries, the number of hospitalizations increased by an estimated 12 percent over a 12-month follow-up period.
- There was no estimated reduction in total Medicare expenditures.

Although the awardee did not explicitly aim to reduce inpatient stays or visits to specialists, there is evidence of short-term statistically significant estimated reductions for both services among Medicaid treatment beneficiaries. The initiative did not sustain its estimated reductions in inpatient stays and specialist visits in the quarter after enrollment over the full year of follow-up. The short-term estimated reduction in specialist visits could reflect a substitution effect in which the CM unit furnished services that an office-based setting would otherwise have delivered. It might also reflect fewer follow-up visits that can sometimes occur after an ED visit.

The documentation challenges encountered by the awardee played a potential role in the impact analysis. The design of the EMS system did not support tracking of billing information. This affected the quality of the insurance information that the system documented and the ability to link beneficiaries to claims and enrollment data. This, in turn, reduced the sample size and ability to detect impacts. And, importantly, it is unknown if the beneficiaries linked to administrative data were representative of the larger set of Medicaid and Medicare beneficiaries eligible to receive a CM visit or who received one.

CONCLUSION

In line with program expectations, there were estimated reductions in the number of ambulance transports for Medicaid and Medicare beneficiaries and ambulance expenditures for Medicare beneficiaries that were sustained over a 12-month follow-up period. There was no effect on total Medicare expenditures. However, there was an estimated increase in the number of ED visits over the full follow-up year for Medicaid beneficiaries. The subsequent increased utilization could reflect the fact the treatment beneficiaries deferred care to a later date and needed more intensive services. Or it could reflect remaining differences between the treatment and matched comparison beneficiaries. The inability to measure the potential role of these factors makes it difficult to fully assess the CCRI's success.

Although the program did not target hospitalizations or specialist visits, there were short-term estimated reductions for these services among Medicaid beneficiaries. CM unit services could have substituted for those furnished by specialists. There is not a clear explanation for the effects on hospitalizations. For Medicare beneficiaries, the number of hospitalizations increased by an estimated 12 percent over a 12-month follow-up period, but the theory of change provides no supporting reason for this observed change in utilization.

The current Medicaid impact estimates reflect the likely impacts over the full target population, with the relatively high participation rate among Medicaid beneficiaries (75 percent) providing a reasonable level of confidence in the generalizability of the findings on service utilization. The program could have diverted other beneficiaries in the treatment group from the ED if the awardee had engaged them. This is especially true for the Medicare population, which had a much lower participation rate (44 percent). In this case, the impact estimates might understate the intervention's potential impacts.

Limitations of evaluation

There is concern about unobservable differences between the treatment and comparison groups that the models could not account for and could lead to misestimation of true program effects in either direction. The treatment and comparison groups were generally well balanced on observable baseline characteristics and the program restricted ambulance transports to those likely associated with a 911 call in the comparison group. However, the inability to classify the reason for dispatch in the comparison group and then match on this introduces concerns about unobservable characteristics that might affect future service use and expenditures. Unobserved social determinants of health might also introduce bias.

PROGRAM SUSTAINABILITY

After its award ended in February 2018, the CCRI stopped operating its program due to lack of funding. A few months before the award ended, the awardee scaled back the CM initiative due to staff attrition, choosing to maintain the medical unit but discontinuing one of the two behavioral health units.

CCRI's proposed payment model

The awardee relied on FFS billing codes to generate revenue from Medicaid and commercial payers to cover the cost of (1) mobile unit treatment without ambulance transport, (2) transport to the ED for patients who could not be treated at home, and (3) clinicians' assessments of 911 callers in behavioral health crisis. The awardee concluded that the only way to sustain the program would be through coverage for intervention services under the Medicare FFS program.

The awardee did not develop a payment model that covered the costs of the program. Although the awardee billed some payers for CM services using existing FFS billing codes, third-party payments covered only about one-quarter of program costs. One challenge to using the FFS payment approach was that many of the core program services did not have a billing code. Another challenge was that the codes could not be used for Medicare because Medicare reimburses only for emergency transportation to a hospital and does not reimburse for on-site care provided by paramedics or APPs.

The awardee pursued other payment models with private payers, but payers lacked sufficient incentives to enter such arrangements. Payers supported the program but were reluctant to increase the treat-no-transport reimbursement to account for APP services or to enter into shared savings agreements. They were not convinced that the additional APP services would divert enough patients from the ED to generate meaningful cost savings, especially given (1) the relatively small number of people affected for any given payer, (2) the lack of evidence of savings, and (3) the lack of federal leadership in reimbursing for on-site low-acuity services. Also, reductions in municipal budgets prevented the local government from dedicating resources to the program. The awardee ultimately concluded that sustaining the program would require Medicare FFS payments for CM unit services.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach with beneficiary fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of outcomes.

In addition to the standard outcomes, impacts were estimated on the number of ambulance transports for Medicaid beneficiaries and ambulance expenditures for Medicare beneficiaries. Ambulance transports were identified using the place of service on the claims and two restrictions were imposed to identify ambulance transports most likely associated with 911 calls. One of the Healthcare Common Procedure Coding System (HCPCS) modifier codes on the claim had to indicate the origin of the ambulance transport was a residential, domiciliary, custodial facility, or a nursing home outside of a skilled nursing facility; residence; or scene of an accident or acute event. Another HCPCS modifier code on the claim had to indicate that a hospital was the destination.

The awardee's finder file included 113,806 unique 911 callers of which 7,015 were linked to Medicaid enrollment data and 5,266 linked to Medicare enrollment data. The impact analysis included 2,872 of the 7,015 Medicaid beneficiaries identified in the awardee's finder file through February 2018 (Table A.1). The study dropped most beneficiaries from the finder file because they lacked 90 days of Medicaid enrollment in the baseline period (35 percent). It excluded another 11 percent because they were not enrolled in Medicaid on the day of enrollment. And it excluded about 9 percent because they did not have a community medicine (CM) visit or an ambulance transport within two days of the 911 call. The study excluded the remaining beneficiaries for other reasons, such as dual eligibility for Medicaid and Medicare or enrollment in the Children's Health Insurance Program.

Table A.1. Identifying the final sample for impact analysis for CCRI: Medicaid

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total number of 911 callers in the awardee's finder file linked to Medicaid enrollment data through February 28, 2018		7,015
Beneficiaries not enrolled in Medicaid on the day of enrollment	763	6,252
Beneficiaries who lacked 90 days of Medicaid enrollment during baseline period	2,518	3,734
Beneficiaries who died, had private insurance, restricted benefits, enrolled in state Children's Health Insurance Program, or were dually eligible for Medicare and Medicaid in the month of enrollment	261	3,473
Beneficiaries who did not receive a CM visit and who did not have an ambulance claim within two days of the 911 call	593	2,880

Table A.1 (*continued*)

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Beneficiaries dropped in matching	8	2,872
Final analytic sample		**2,872**

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018.

CCRI = Community Care Response Initiative; CM = community medicine.

The impact analysis included 1,750 Medicare beneficiaries among the 5,266 beneficiaries identified in the awardee's finder file through February 2018 (Table A.2). The study excluded most beneficiaries (47 percent) from the analytic sample because they were enrolled in Medicare Advantage. It excluded about 8 percent because they did not have a CM visit or an ambulance transport within two days of the 911 call or it dropped them during matching. The study excluded the remaining beneficiaries for reasons such as lacking 90 days of FFS enrollment in the baseline period or Medicare not being the primary payer.

Table A.2. Identifying the final sample for impact analysis for CCRI: Medicare

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total number of 911 callers in the awardee's finder file linked to Medicare enrollment data through February 28, 2018		5,266
Beneficiaries who did not meet the standard claims-based inclusion criteria		
Not enrolled in both Medicare Parts A and B	576	4,690
Enrolled in Medicare Advantage	2,453	2,237
Medicare was not the primary payer	29	2,208
Lacked 90 days of FFS enrollment during baseline period	63	2,145
Beneficiaries who did not receive a CM visit and who did not have an ambulance claim within two days of the 911 call	372	1,773
Beneficiaries dropped in matching	23	1,750
Final analytic sample		1,750

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018.

CCRI = Community Care Response Initiative; CM = community medicine; FFS = fee-for-service.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching for the Medicaid sample. The table displays the weighted means of baseline characteristics for the 2,872 treatment beneficiaries and the 11,291 matched comparison beneficiaries used in the Medicaid impact analysis. The matching variables include demographic characteristics; Medicaid eligibility; health status (as measured by the Chronic Illness and Disability Payment System [CDPS] score and chronic condition indicators); and service use. The analysis required exact matches on the following variables: enrollment period for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, CDPS score quartile, Medicaid enrollment in the year before baseline, and behavioral health encounter.

Table B.2 presents the baseline characteristics of the 1,750 Medicare treatment beneficiaries and 6,014 matched comparison beneficiaries used in the Medicare impact analysis. The matching variables include demographic characteristics (age, gender, and race); Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category [HCC] score and chronic condition indicators); Medicare expenditures in total and by type of service; and service use. Exact matches were required for the enrollment period for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, HCC score above the 50th percentile, fee-for-service enrollment in the year before baseline, and behavioral health encounter.

The tables show the means, difference in means, the percentage difference, and the standardized difference for each variable, which the analysis calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. However, some of the percentage differences for use measures were more than 10 percent, which is nontrivial relative to the size of potential impacts. This, and other larger percentage differences, might suggest some unobserved remaining bias between the treatment and comparison groups.

The tables also show the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence of the statistical significance of the difference in the means. The equivalence test *p*-values are the greater of the *p*-values for two one-sided tests of whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the analysis included an omnibus test in which the null hypothesis is that the treatment and matched comparison groups balanced across all linear combinations of the covariates. The results assess the closeness of fit between the treatment and matched comparison groups on characteristics likely associated with outcomes. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

Tables B.3 and B.4 shows the differences between 911 callers receiving a community medicine (CM) visit and those transported to the emergency department (ED) via ambulance in the Medicaid and Medicare samples, respectively. This evidence supports an intent-to-treat analysis based on all 911 callers rather than only those who received a CM visit. Medicaid beneficiaries

who received a CM visit had, on average, lower risk scores and lower levels of baseline service use and disease burden compared to those transported by an ambulance to the ED. Conversely, Medicare beneficiaries who received a CM visit had, on average, higher risk scores, greater baseline expenditures and service use, and were more likely to be dually eligible for Medicare and Medicaid. Given the CCRI's intent was to dispatch the CM units for low-acuity cases, it does not appear that the dispatchers were successful at doing so among Medicare beneficiaries. The factors behind the differences are likely due to characteristics that cannot be fully identified in claims, such as the reason for the 911 call and social determinants of health.

Table B.1. Baseline characteristics of treatment and matched beneficiaries for CCRI: Medicaid sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	35 (0.26)	33 (0.16)	2.6 (0.41)	7.3	0.18	< 0.01	< 0.01
Male, %	48 (0.93)	49 (0.47)	-0.67 (1.3)	-1.4	-0.01	0.60	< 0.01
Medicaid enrollment							
Number of Medicaid-eligible days	315 (1.6)	318 (0.69)	-3.3 (2.2)	-1.0	-0.04	0.13	< 0.01
Number of Medicaid FFS-eligible days	53 (1.9)	48 (0.99)	4.9 (2.7)	9.3	0.05	0.07	< 0.01
Number of days enrolled in FFS in last quarter of the baseline year	12 (0.54)	12 (0.27)	0.61 (0.74)	4.9	0.02	0.41	< 0.01
Number of Medicaid-eligible days in 4th quarter before enrollment	66 (0.74)	67 (0.34)	-0.74 (1.1)	-1.1	-0.02	0.48	< 0.01
Number of Medicaid-eligible days in 3rd quarter before enrollment	74 (0.62)	75 (0.28)	-1.4 (0.86)	-1.9	-0.04	0.11	< 0.01
Number of Medicaid-eligible days in 2nd quarter before enrollment	84 (0.36)	85 (0.15)	-1.2 (0.49)	-1.4	-0.06	0.01	< 0.01
Number of Medicaid-eligible days in quarter before enrollment	91 (0.07)	91 (0.03)	0.06 (0.11)	< +/-1	0.02	0.56	< 0.01
Health status and diagnoses							
CDPS score[a]	2.2 (0.04)	2.2 (0.02)	0.02 (0.06)	1.1	0.01	0.65	< 0.01
AIDS or other infectious disease, %	10 (0.57)	11 (0.29)	-0.77 (0.81)	-7.4	-0.03	0.35	< 0.01
Cardiovascular disease, %	31 (0.87)	27 (0.41)	4.0 (1.2)	13	0.09	< 0.01	< 0.01
Central nervous system condition, %	17 (0.70)	15 (0.33)	1.7 (0.92)	10	0.05	0.06	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Diabetes, %	13 (0.63)	13 (0.32)	-0.16 (0.87)	-1.2	0.00	0.85	< 0.01
Disabled, %	15 (0.67)	19 (0.36)	-3.3 (0.99)	-21	-0.09	< 0.01	< 0.01
Gastrointestinal condition, %	23 (0.79)	23 (0.38)	0.28 (1.2)	1.2	0.01	0.81	< 0.01
Genital condition, %	5.8 (0.44)	5.4 (0.21)	0.47 (0.60)	8.1	0.02	0.43	< 0.01
Metabolic condition, %	13 (0.64)	15 (0.31)	-1.1 (0.92)	-8.1	-0.03	0.24	< 0.01
Psychiatric condition, %	44 (0.93)	45 (0.46)	-0.80 (1.3)	-1.8	-0.02	0.54	< 0.01
Pulmonary condition, %	27 (0.83)	26 (0.40)	0.71 (1.1)	2.7	0.02	0.54	< 0.01
Renal condition, %	7.8 (0.50)	7.0 (0.23)	0.82 (0.67)	11	0.03	0.22	< 0.01
Skeletal condition, %	27 (0.82)	23 (0.39)	3.4 (1.1)	13	0.08	< 0.01	< 0.01
Skin condition, %	18 (0.72)	16 (0.33)	1.9 (0.99)	11	0.05	0.05	< 0.01
Substance abuse, %	37 (0.90)	32 (0.43)	5.0 (1.2)	14	0.10	< 0.01	< 0.01
Service use							
Total hospitalizations	782 (33)	744 (12)	39 (45)	4.9	0.02	0.39	< 0.01
Number of ambulance transports in preprogram period imputation	1,061 (54)	929 (15)	132 (72)	12	0.05	0.07	< 0.01
Number of ED or observation visits in preprogram period imputation	2.0 (0.08)	1.8 (0.02)	0.20 (0.11)	9.8	0.05	0.06	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Primary care visits, ambulatory setting	4,083 (123)	3,666 (51)	417 (165)	10	0.06	0.01	< 0.01
Specialist visits, ambulatory setting	5,334 (163)	4,756 (69)	578 (212)	11	0.07	< 0.01	< 0.01
Total 30-day unplanned readmissions	101 (11)	81 (3.5)	20 (14)	20	0.04	0.16	< 0.01
Area-level factors							
Median household income in zip code of residence	43,635 (208)	39,343 (103)	4,292 (314)	9.8	0.38	< 0.01	1.00
Percentage with private health insurance in zip code of residence	54 (0.22)	51 (0.12)	2.6 (0.34)	4.8	0.22	< 0.01	0.13
Percentage with public health insurance in zip code of residence	34 (0.15)	40 (0.06)	-5.9 (0.20)	-17	-0.74	< 0.01	1.00
Percentage of adults ages 25 or older in the county with a degree from a 4-year college	30 (0.06)	27 (0.06)	2.7 (0.12)	9.2	0.87	< 0.01	1.00
Percentage Hispanic in zip code of residence	32 (0.25)	44 (0.21)	-12 (0.49)	-37	-0.88	< 0.01	1.00
Percentage White in zip code of residence	80 (0.21)	76 (0.12)	4.1 (0.31)	5.1	0.37	< 0.01	1.00
Propensity score	0.75 (0.00)	0.75 (0.00)	0.00 (0.00)	< +/-1	0.00	0.96	< 0.01
Number of beneficiaries	2,872	11,291					
Omnibus test				Chi-squared statistic	Degrees of freedom	*p*-value	
				11,852.38	78.00	0.00	

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research-identifiable files or other data sources.

Note: Standard errors in parentheses. Standardized difference is calculated as the ratio of the adjusted difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values. The comparison

Table B.1 (*continued*)

group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid. Exact matching variables include FFS enrollment in the year before baseline, CDPS score above the 50th percentile, behavioral health encounter, semi-annual period of enrollment, and program year.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

CCRI = Community Care Response Initiative; CDPS = Chronic Illness and Disability Payment System; ED = emergency department; FFS = fee-for-service; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System; SE = standard error.

Table B.2. Baseline characteristics of treatment and matched beneficiaries for CCRI: Medicare sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics, %							
Male	46 (1.2)	47 (0.64)	-1.1 (1.7)	-2.4	-0.02	0.52	< 0.01
Female	54 (1.2)	53 (0.64)	1.1 (1.7)	2.1	0.02	0.52	< 0.01
Age: younger than 65	21 (0.97)	18 (0.40)	3.1 (1.3)	15	0.08	0.02	< 0.01
Age: 65 to 74	29 (1.1)	29 (0.59)	0.08 (1.5)	< +/-1	0.00	0.96	< 0.01
Age: 75 to 84	28 (1.1)	31 (0.62)	-3.3 (1.6)	-12	-0.07	0.04	< 0.01
Age: 85 and older	22 (0.99)	22 (0.54)	0.07 (1.5)	< +/-1	0.00	0.96	< 0.01
White	91 (0.68)	91 (0.33)	0.19 (0.96)	< +/-1	0.01	0.85	< 0.01
Black	3.3 (0.42)	1.5 (0.13)	1.7 (0.51)	52	0.10	< 0.01	< 0.01
Hispanic	2.0 (0.33)	2.4 (0.16)	-0.39 (0.51)	-20	-0.03	0.44	< 0.01
Other	3.6 (0.45)	5.1 (0.26)	-1.5 (0.68)	-42	-0.08	0.03	< 0.01
Medicare entitlement and dual eligibility status, %							
Original reason for Medicare entitlement: age	68 (1.1)	72 (0.52)	-4.1 (1.5)	-6.0	-0.09	< 0.01	< 0.01
Original reason for Medicare entitlement: disability	30 (1.1)	26 (0.51)	3.5 (1.5)	12	0.08	0.02	< 0.01
Original reason for Medicare entitlement: ESRD	2.1 (0.34)	1.5 (0.12)	0.58 (0.45)	28	0.04	0.19	< 0.01
Dually eligible for Medicare and Medicaid	23 (1.0)	21 (0.47)	1.5 (1.3)	6.7	0.04	0.26	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Health status and diagnoses							
HCC score[a]	2.09 (0.04)	1.99 (0.02)	0.10 (0.06)	4.7	0.06	0.10	< 0.01
Acute renal failure, %	12 (0.79)	12 (0.40)	0.66 (1.1)	5.3	0.02	0.56	< 0.01
Cardio-respiratory failure and shock, %	11 (0.75)	12 (0.40)	-1.1 (1.1)	-9.8	-0.03	0.33	< 0.01
CHF, %	24 (1.0)	23 (0.54)	1.2 (1.4)	4.8	0.03	0.42	< 0.01
Coagulation defects and other hematological disorders, %	11 (0.76)	9.9 (0.37)	1.4 (1.0)	12	0.04	0.17	< 0.01
COPD, %	26 (1.0)	25 (0.55)	0.05 (1.5)	< +/-1	0.00	0.97	< 0.01
Diabetes with chronic complications, %	21 (0.97)	16 (0.47)	4.3 (1.3)	21	0.11	< 0.01	< 0.01
Diabetes without complication, %	11 (0.76)	12 (0.42)	-1.0 (1.1)	-9.1	-0.03	0.35	< 0.01
Major depressive, bipolar, and paranoid disorders, %	15 (0.86)	11 (0.37)	3.8 (1.2)	25	0.11	< 0.01	< 0.01
Protein-calorie malnutrition, %	6.5 (0.59)	6.0 (0.27)	0.49 (0.82)	7.5	0.02	0.55	< 0.01
Rheumatoid arthritis and inflammatory connective tissue disease, %	11 (0.74)	8.8 (0.37)	2.0 (1.0)	18	0.06	0.06	< 0.01
Septicemia, sepsis, inflammatory response syndrome/shock	9.2 (0.33)	9.2 (0.69)	0.00 (0.95)	< +/-1	0.00	1.00	< 0.01
Vascular disease	21 (0.52)	25 (1.0)	4.4 (1.4)	18	0.10	< 0.01	< 0.01
Specified heart arrhythmias, %	27 (1.1)	25 (0.56)	2.3 (1.5)	8.6	0.05	0.11	< 0.01
Medicare expenditures							
Total expenditures	2,210 (81)	2,028 (36)	182 (115)	8.2	0.05	0.11	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total expenditures, 3 months before enrollment	2,745 (128)	2,675 (66)	70 (190)	2.6	0.01	0.71	< 0.01
Outpatient expenditures	299 (16)	305 (8.0)	-5.8 (23)	-1.9	-0.01	0.80	< 0.01
Physician services expenditures	564 (18)	459 (8.5)	105 (26)	19	0.14	< 0.01	< 0.01
Ambulance expenditures	19 (1.4)	22 (0.64)	-3.0 (2.1)	-16	-0.05	0.16	< 0.01
Service use							
Total hospitalizations	806 (37)	724 (15)	83 (50)	10	0.05	0.10	< 0.01
Total hospitalizations, 3 months before enrollment	1,071 (63)	919 (26)	152 (83)	14	0.06	0.07	< 0.01
Total emergency ambulance transports	660 (48)	656 (20)	4.2 (71)	< +/-1	0.00	0.95	< 0.01
Primary care visits, ambulatory setting	7,909 (201)	6,702 (85)	1,207 (260)	15	0.14	< 0.01	< 0.01
Primary care visits, ambulatory setting, 3 months before enrollment	9,108 (308)	7,653 (130)	1,454 (393)	16	0.11	< 0.01	< 0.01
Outpatient ED or observation visits	1,368 (91)	1,277 (33)	91 (127)	6.7	0.02	0.47	< 0.01
Specialist visits, any setting	19,273 (546)	14,024 (181)	5,249 (675)	27	0.23	< 0.01	0.25
Area-level factors							
No part of county of residence designated HPSA, %	1.5 (0.29)	1.5 (0.15)	0.08 (0.43)	5.2	0.01	0.85	< 0.01
Entire county of residence designated HPSA, %	1.1 (0.25)	1.5 (0.16)	-0.38 (0.38)	-35	-0.04	0.32	< 0.01
One or more parts of county of residence designated HPSA, %	97 (0.38)	97 (0.22)	0.30 (0.57)	< +/-1	0.02	0.60	< 0.01
Median household income in zip code	49,745 (333)	49,720 (204)	25 (515)	< +/-1	0.00	0.96	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Adults in the county with four-year college degree, %	29 (0.12)	28 (0.08)	1.3 (0.20)	4.5	0.25	< 0.01	0.54
Percentage with public health insurance in zip code	35 (0.23)	38 (0.13)	-2.9 (0.34)	-8.3	-0.30	< 0.01	0.94
Percentage with private health insurance in zip code	63 (0.31)	62 (0.17)	0.91 (0.45)	1.4	0.07	0.04	< 0.01
Propensity score	0.44 (0.01)	0.44 (0.00)	0.00 (0.01)	< +/-1	0.00	0.88	< 0.01
Number of beneficiaries	1,750	6,014					
Omnibus test				Chi-squared statistic 1103.83	Degrees of freedom 44.00	*p*-value 0.00	

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicare claims and enrollment data from December 2012 through August 2018.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the adjusted difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include FFS enrollment in the year before baseline, HCC score above the 50th percentile, behavioral health encounter, semi-annual period of enrollment, and program year.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CCRI = Community Care Response Initiative; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disorder; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; HPSA = health professional shortage area; SE = standard error.

Table B.3. Baseline characteristics of 911 callers who received a CM visit and those transported by an ambulance to the ED: Medicaid sample

Characteristic	CM visit mean (SE)	911 transport mean (SE)	Adjusted difference	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	34 (0.14)	38 (0.31)	-3.9	-11	-0.27	0.00	0.70
Male, %	0.48 (0.01)	0.49 (0.01)	-0.01	-3	-0.03	0.51	0.00
Health status and diagnoses							
CDPS score[a]	2.1 (0.03)	2.6 (0.04)	-0.52	-25	-0.27	0.00	0.66
First quartile of the CDPS score, %	0.27 (0.00)	0.19 (0.01)	0.08	31	0.19	0.00	0.06
Second quartile of the CDPS score, %	0.25 (0.00)	0.25 (0.01)	0.00	1.5	0.01	0.84	0.00
Third quartile of the CDPS score, %	0.24 (0.00)	0.27 (0.01)	-0.03	-12	-0.07	0.10	0.00
Fourth quartile of the CDPS score, %	0.23 (0.00)	0.29 (0.01)	-0.06	-25	-0.14	0.00	0.01
AIDS or other infectious disease, %	0.10 (0.00)	0.12 (0.01)	-0.02	-16	-0.05	0.25	0.00
Cardiovascular disease, %	0.29 (0.01)	0.40 (0.01)	-0.11	-38	-0.24	0.00	0.42
Central nervous system condition, %	0.15 (0.00)	0.22 (0.01)	-0.07	-43	-0.18	0.00	0.09
Diabetes, %	0.12 (0.00)	0.16 (0.01)	-0.04	-30	-0.11	0.02	0.00
Disabled, %	0.15 (0.00)	0.16 (0.01)	-0.01	-5.2	-0.02	0.62	0.00
Gastrointestinal condition, %	0.22 (0.00)	0.27 (0.01)	-0.06	-26	-0.14	0.00	0.01
Genital condition, %	0.06 (0.00)	0.06 (0.01)	0.00	6.3	0.02	0.71	0.00
Metabolic condition, %	0.13 (0.00)	0.16 (0.01)	-0.03	-22	-0.09	0.07	0.00

Table B.3 (*continued*)

Characteristic	CM visit mean (SE)	911 transport mean (SE)	Adjusted difference	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Psychiatric condition, %	0.45 (0.01)	0.43 (0.01)	0.02	4.4	0.04	0.36	0.00
Pulmonary condition, %	0.25 (0.01)	0.31 (0.01)	-0.06	-23	-0.13	0.00	0.01
Renal condition, %	0.07 (0.00)	0.12 (0.01)	-0.05	-77	-0.2	0.00	0.17
Skeletal condition, %	0.26 (0.00)	0.28 (0.01)	-0.02	-9.3	-0.06	0.21	0.00
Skin condition, %	0.18 (0.00)	0.17 (0.01)	0.01	6.1	0.03	0.50	0.00
Substance abuse, %	0.37 (0.01)	0.38 (0.01)	-0.01	-3.3	-0.03	0.57	0.00
Service use							
Total hospitalizations	688 (25)	1,066 (32)	-378	-55	-0.25	0.00	0.52
Number of ambulance transports in preprogram period imputation	936 (43)	1,436 (52)	-500	-53	-0.21	0.00	0.25
Number of ED or observation visits in preprogram period imputation	1.9 (0.05)	2.2 (0.09)	-0.30	-15	-0.08	0.13	0.00
Primary care visits, ambulatory setting	3,960 (71)	4,452 (142)	-492	-12	-0.08	0.08	0.00
Specialist visits, ambulatory setting	5,003 (107)	6,327 (179)	-1,32	-26	-0.16	0.00	0.03
Total 30-day unplanned readmissions	67 (9.3)	203 (9.1)	-135	-201	-0.32	0.00	0.81
Behavioral health mobile unit encounter, %	0.22 (0.00)	0.09 (0.01)	0.13	60	0.31	0.00	0.97
Propensity score	0.77 (0.00)	0.69 (0.00)	0.09	11	0.8	0.00	1.00
Number of beneficiaries	2,154	718					

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research-identifiable files or other data sources.

Table B.3 (*continued*)

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the adjusted difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; FFS = fee-for-service; SE = standard error; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System.

Table B.4. Baseline characteristics of 911 callers who received a CM visit and those transported by an ambulance to the ED: Medicare sample

Characteristic	CM visit mean (SE)	911 transport mean (SE)	Adjusted difference	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics, %							
Male	45 (1.8)	47 (1.6)	-2.8	-6.4	-0.06	0.23	< 0.01
Female	55 (1.8)	53 (1.6)	2.8	5.1	0.06	0.23	< 0.01
Age: younger than 65	31 (1.7)	13 (1.1)	19	59	0.40	< 0.01	1.00
Age: 65 to 74	28 (1.6)	30 (1.5)	-1.5	-5.3	-0.03	0.48	< 0.01
Age: 75 to 84	22 (1.5)	33 (1.5)	-11	-51	-0.27	< 0.01	0.65
Age: 85 and older	19 (1.4)	25 (1.4)	-6.0	-33	-0.16	< 0.01	0.03
White	89 (1.1)	93 (0.81)	-4.4	-5.0	-0.14	< 0.01	< 0.01
Black	4.5 (0.74)	2.3 (0.48)	2.2	50	0.11	0.01	< 0.01
Hispanic	2.8 (0.60)	1.3 (0.37)	1.5	53	0.09	0.03	< 0.01
Other	4.0 (0.70)	3.3 (0.57)	0.70	18	0.04	0.44	< 0.01
Medicare entitlement and dual eligibility status, %							
Original reason for Medicare entitlement: age	56 (1.8)	77 (1.3)	-21	-37	-0.42	< 0.01	1.00
Original reason for Medicare entitlement: disability	41 (1.8)	21 (1.3)	20	49	0.41	< 0.01	1.00
Original reason for Medicare entitlement: ESRD	2.6 (0.57)	1.6 (0.41)	0.93	36	0.06	0.19	< 0.01
Dually eligible for Medicare and Medicaid	32 (1.7)	15 (1.1)	17	54	0.37	< 0.01	1.00
FFS in baseline	93 (0.94)	95 (0.67)	-2.7	-2.9	-0.10	0.02	< 0.01

Table B.4 (*continued*)

Characteristic	CM visit mean (SE)	911 transport mean (SE)	Adjusted difference	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Health status and diagnoses							
HCC score[a]	2.21 (0.07)	2.00 (0.05)	0.21	9.6	0.11	0.01	< 0.01
HCC score above 50th percentile, %	48 (1.8)	46 (1.6)	1.7	3.5	0.03	0.48	< 0.01
Acute renal failure, %	13 (1.2)	12 (1.0)	1.2	9.2	0.04	0.44	< 0.01
Cardio-respiratory failure and shock, %	11 (1.1)	11 (1.00)	0.66	5.8	0.02	0.67	< 0.01
CHF, %	23 (1.5)	25 (1.4)	-1.9	-7.9	-0.04	0.37	< 0.01
Coagulation defects and other hematological disorders, %	12 (1.1)	11 (1.0)	0.48	4.2	0.02	0.74	< 0.01
COPD, %	27 (1.6)	24 (1.4)	2.9	11	0.07	0.17	< 0.01
Diabetes with chronic complications, %	20 (1.4)	21 (1.3)	-1.7	-8.6	-0.04	0.39	< 0.01
Diabetes without complication, %	11 (1.1)	12 (1.0)	-0.19	-1.6	-0.01	0.90	< 0.01
Major depressive, bipolar, and paranoid disorders, %	18 (1.4)	13 (1.1)	5.6	30	0.14	< 0.01	0.01
Protein-calorie malnutrition, %	8.0 (0.97)	5.3 (0.72)	2.6	33	0.10	0.03	< 0.01
Rheumatoid arthritis and inflammatory connective tissue disease, %	12 (1.2)	10.0 (0.96)	1.9	16	0.06	0.23	< 0.01
Septicemia, sepsis, inflammatory response syndrome, or shock, %	13 (1.2)	6.4 (0.78)	6.4	50	0.19	< 0.01	0.09
Vascular disease, %	23 (1.5)	27 (1.4)	-4.2	-18	-0.10	0.04	< 0.01
Specified heart arrhythmias, %	24 (1.5)	29 (1.5)	-5.0	-20	-0.12	0.02	< 0.01

Table B.4 (*continued*)

Characteristic	CM visit mean (SE)	911 transport mean (SE)	Adjusted difference	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Medicare expenditures							
Total expenditures	2,478 (128)	1,997 (104)	482	19	0.14	< 0.01	< 0.01
Total expenditures, 3 months before enrollment	3,132 (212)	2,436 (156)	696	22	0.12	< 0.01	< 0.01
Outpatient expenditures	330 (24)	275 (21)	56	17	0.08	0.08	< 0.01
Physician services expenditures	598 (26)	536 (24)	62	10	0.09	0.08	< 0.01
Ambulance expenditures	28 (2.6)	11 (1.3)	17	60	0.23	< 0.01	0.30
Service use							
Total hospitalizations	943 (59)	697 (46)	245	26	0.15	< 0.01	0.01
Total hospitalizations, 3 months before enrollment	1,230 (99)	944 (82)	286	23	0.10	0.03	< 0.01
Outpatient ED or observation visits	1,895 (181)	947 (74)	947	50	0.19	< 0.01	0.05
Total emergency ambulance transports	982 (91)	402 (47)	580	59	0.23	< 0.01	0.31
Primary care visits, ambulatory setting	7,858 (301)	7,950 (269)	-93	-1.2	-0.01	0.81	< 0.01
Primary care visits, ambulatory setting, 3 months before enrollment	8,885 (446)	9,285 (426)	-400	-4.5	-0.03	0.52	< 0.01
Specialist visits, any setting	21,205 (859)	17,730 (700)	3,475	16	0.15	< 0.01	0.01
Behavioral encounter, %	11 (1.1)	1.2 (0.35)	10.0	89	0.32	< 0.01	0.96
Propensity score	0.58 (0.01)	0.33 (0.01)	0.25	42	1.09	< 0.01	1.00
Number of beneficiaries	777	973					

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicare claims and enrollment data from December 2012 through August 2018.

Table B.4 (*continued*)

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the adjusted difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CHF = congestive heart failure; COPD = chronic obstructive pulmonary disorder; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; SE = standard error.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Table C.1 displays the results from the impact analysis for the Medicaid population; Table C.2 shows results for the Medicare population. Table C.3 shows the impact estimates for the non-top-coded Medicaid utilization outcomes. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the 0.10, 0.05, and 0.01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the CCRI on select Medicaid utilization measures during a 12-month follow-up period

	All eligible beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED visits or observation stays, per 1,000 beneficiaries[b]					
Baseline year	4,266	3,849			
Months 1–3	7,951	7,774	-241 (180)	-3.0%	0.18
Months 4–6	4,805	3,680	708*** (164)	18%	< 0.01
Months 7–9	4,541	3,320	803*** (172)	23%	< 0.01
Months 10–12	4,496	3,265	814*** (180)	23%	< 0.01
Months 1–12	5,611	4,640	553*** (139)	11%	< 0.01
Number of ambulance transports, per 1,000 beneficiaries[b]					
Baseline year	1,191	1,089			
Months 1–3	3,930	5,648	-1,821*** (86)	-32%	< 0.01
Months 4–6	1,621	1,352	168** (76)	12%	0.03
Months 7–9	1,463	1,249	112 (75)	9.3%	0.13
Months 10–12	1,548	1,196	250*** (79)	22%	< 0.01
Months 1–12	2,272	2,536	-366*** (63)	-14%	< 0.01
Hospital stays, per 1,000 beneficiaries[b]					
Baseline year	722	658			
Months 1–3	1,564	1,775	-276*** (54)	-15%	< 0.01
Months 4–6	842	682	96** (44)	14%	0.03
Months 7–9	767	631	71 (43)	11%	0.10
Months 10–12	827	560	202*** (45)	38%	< 0.01
Months 1–12	1,048	974	8.7 (36)	< 1%	0.81
Primary care visits in ambulatory settings, per 1,000 beneficiaries[b]					
Baseline year	3,953	3,519			
Months 1–3	5,246	4,966	-154 (140)	-3.0%	0.27
Months 4–6	4,269	3,749	86 (135)	2.2%	0.52
Months 7–9	4,260	3,571	254* (142)	6.7%	0.07
Months 10–12	4,229	3,447	348** (150)	9.5%	0.02
Months 1–12	4,571	4,037	100 (117)	2.3%	0.39

Table C.1 (*continued*)

	All eligible beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Specialist visits in ambulatory settings, per 1,000 beneficiaries[b]					
Baseline year	5,289	4,657			
Months 1–3	7,836	7,910	-707*** (197)	-8.5%	< 0.01
Months 4–6	6,412	5,608	172 (190)	2.9%	0.37
Months 7–9	6,164	5,246	285 (197)	5.1%	0.15
Months 10–12	6,326	5,218	475** (217)	8.5%	0.03
Months 1–12	6,833	6,155	46 (161)	< 1%	0.78
Beneficiaries with a readmission, %					
Baseline year	5.5	4.6			
Months 1–3	4.8	5.6	-0.86 (0.73)	-15%	0.24
Months 4–6	2.3	3.1	-0.68 (0.68)	-22%	0.32
Months 7–9	1.9	2.8	-0.87 (0.68)	-32%	0.20
Months 10–12	2.4	2.3	0.10 (0.71)	4.6%	0.88
Months 1–12	8.7	9.1	-0.36 (1.6)	-4.0%	0.82
Sample sizes					
Number of beneficiaries					
Baseline year	2,872	11,291			
Months 1–3	2,872	11,291			
Months 4–6	2,781	10,931			
Months 7–9	2,611	10,326			
Months 10–12	2,448	9,621			
Months 1–12	2,872	11,291			

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research-identifiable files or other data sources.

Notes: Impacts on expenditures are not estimated because of the high penetration of managed care in Arizona.

Impact estimates for the number of visits, transports, or admissions relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiaries' characteristics and whether they had any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CCRI = Community Care Response Initiative; ED = emergency department; SE = standard error; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System.

Table C.2. Estimated impact of the CCRI on select Medicare FFS expenditures (dollars PBPM) and service utilization measures during a 12-month follow-up period

	All eligible beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)					
Baseline year	2,184	2,009			
Months 1–3	6,974	6,892	-93 (283)	-1.3%	0.74
Months 4–6	3,223	2,727	321** (160)	13%	0.05
Months 7–9	2,999	2,544	281* (170)	13%	0.10
Months 10–12	3,086	2,612	299* (163)	13%	0.07
Months 1–12	4,234	3,885	174 (142)	4.8%	0.22
Ambulance expenditures ($ PBPM)					
Baseline year	28	29			
Months 1–3	158	208	-49*** (4.2)	-24%	< 0.01
Months 4–6	49	42	8.0** (3.5)	21%	0.02
Months 7–9	44	40	4.6 (3.6)	13%	0.20
Months 10–12	42	39	3.0 (3.4)	9.1%	0.37
Months 1–12	79	91	-11*** (2.6)	-13%	< 0.01
ED visits or observation stays, per 1,000 beneficiaries					
Baseline year	1,883	1,606			
Months 1–3	4,750	4,682	-210 (179)	-4.2%	0.24
Months 4–6	2,165	1,801	87 (144)	4.2%	0.55
Months 7–9	2,148	1,636	234 (148)	12%	0.11
Months 10–12	1,850	1,539	34 (146)	1.9%	0.81
Months 1–12	2,854	2,561	15 (106)	< 1%	0.88
Number of ambulance transports, per 1,000 beneficiaries					
Baseline year	989	857			
Months 1–3	5,505	5,800	-426*** (138)	-7.3%	< 0.01
Months 4–6	1,711	1,211	368*** (116)	30%	< 0.01
Months 7–9	1,521	1,152	237** (117)	21%	0.04
Months 10–12	1,463	1,123	208* (112)	20%	0.06
Months 1–12	2,751	2,561	58 (87)	2.3%	0.51
Hospital stays, per 1,000 beneficiaries					
Baseline year	788	709			
Months 1–3	2,988	2,833	76 (107)	2.7%	0.48
Months 4–6	1,200	928	193*** (74)	23%	< 0.01
Months 7–9	1,064	797	188*** (72)	30%	< 0.01
Months 10–12	1,166	865	221*** (76)	33%	< 0.01
Months 1–12	1,691	1,454	158*** (57)	12%	< 0.01
Primary care visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	7,932	6,716			
Months 1–3	13,409	11,673	520 (476)	4.1%	0.27
Months 4–6	9,507	8,063	227 (386)	2.5%	0.56

Table C.2 (*continued*)

	All eligible beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 7–9	8,616	7,268	131 (331)	1.7%	0.69
Months 10–12	9,162	7,391	554 (380)	7.0%	0.14
Months 1–12	10,357	8,791	349 (286)	3.7%	0.22
Specialist visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	11,307	9,047			
Months 1–3	16,234	13,341	633 (387)	4.1%	0.10
Months 4–6	13,463	10,636	567 (399)	4.4%	0.16
Months 7–9	12,026	9,709	57 (390)	< 1%	0.88
Months 10–12	12,043	9,365	418 (410)	3.7%	0.31
Months 1–12	13,593	10,909	425 (291)	3.3%	0.14
Beneficiaries with a readmission, %					
Baseline year	8.8	7.1			
Months 1–3	14	11	2.3** (1.1)	20%	0.03
Months 4–6	4.2	4.0	0.20 (0.84)	4.9%	0.82
Months 7–9	3.6	2.1	1.5* (0.76)	70%	0.05
Months 10–12	3.4	2.1	1.2 (0.78)	58%	0.12
Months 1–12	21	17	4.0** (1.9)	23%	0.03
Sample sizes					
Number of beneficiaries					
Baseline year	1,750	6,014			
Months 1–3	1,750	6,014			
Months 4–6	1,550	5,154			
Months 7–9	1,420	4,722			
Months 9–12	1,318	4,308			
Months 1–12	1,750	6,014			

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicare claims and enrollment data from December 2012 through August 2018.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiaries' characteristics and whether they had any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CCRI = Community Care Response Initiative; FFS = fee for service; ED = emergency department; PBPM = per beneficiary per month; SE = standard error.

Table C.3. Estimated impact of the CCRI on select non-top-coded Medicaid utilization measures during a 12-month follow-up period

	All eligible beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED visits or observation visits, per 1,000 beneficiaries					
Baseline year	4,376	3,927			
Months 1–3	8,526	8,015	61 (236)	< 1%	0.80
Months 4–6	5,081	3,801	831*** (209)	19%	< 0.01
Months 7–9	4,932	3,378	1,104*** (230)	29%	< 0.01
Months 10–12	4,816	3,366	1,000*** (246)	27%	< 0.01
Months 1–12	5,952	4,787	715*** (171)	14%	< 0.01
Number of ambulance transports, per 1,000 beneficiaries					
Baseline year	1,238	1,151			
Months 1–3	4,412	5,986	-1,661*** (132)	-27%	< 0.01
Months 4–6	1,830	1,509	235** (110)	15%	0.03
Months 7–9	1,692	1,321	285*** (108)	21%	< 0.01
Months 10–12	1,672	1,307	278** (108)	22%	0.01
Months 1–12	2,488	2,676	-274*** (81)	-10%	< 0.01
Hospital stays, per 1,000 beneficiaries					
Baseline year	743	675			
Months 1–3	1,740	1,936	-264*** (70)	-13%	< 0.01
Months 4–6	871	699	105** (53)	14%	0.05
Months 7–9	828	650	111** (54)	16%	0.04
Months 10–12	922	568	287*** (60)	49%	< 0.01
Months 1–12	1,115	1,006	42 (41)	4.0%	0.30
Primary care visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	4,070	3,595			
Months 1–3	5,543	5,471	-403* (206)	-6.8%	0.05
Months 4–6	4,484	3,935	74 (196)	1.7%	0.71
Months 7–9	4,417	3,765	177 (191)	4.3%	0.35
Months 10–12	4,468	3,561	432** (207)	11%	0.04
Months 1–12	4,763	4,245	43 (147)	< 1%	0.77
Specialist visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	5,404	4,733			
Months 1–3	8,360	8,239	-550** (250)	-6.2%	0.03
Months 4–6	6,779	5,845	264 (247)	4.1%	0.28
Months 7–9	6,521	5,562	288 (258)	4.7%	0.27
Months 10–12	6,828	5,466	691** (287)	11%	0.02
Months 1–12	7,174	6,368	135 (195)	2.0%	0.49

Table C.3 (*continued*)

	All eligible beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes					
Number of beneficiaries					
Baseline year	2,872	11,291			
Months 1–3	2,872	11,291			
Months 4–6	2,781	10,931			
Months 7–9	2,611	10,326			
Months 10–12	2,448	9,621			
Months 1–12	2,872	11,291			

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research-identifiable files or other data sources.

Note: Impact estimates for the number of visits, transports, or admissions relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiaries' characteristics and whether they had any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CCRI = Community Care Response Initiative; ED = emergency department; SE = standard error; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the study also estimated program impacts for the Community Care Response Initiative (CCRI) using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. Drawing probabilistic conclusions requires external or prior evidence. In this analysis, the findings from the evaluation of 87 awardees included in Round 1 of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to the CCRI. The study calculated probabilities using the results of a Bayesian regression that jointly models impacts for both Medicare and Medicaid populations on the four core outcomes used by the Centers for Medicare & Medicaid Services (CMS), thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for CMS's four core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. As in the report, these results show impacts on total Medicare expenditures and Medicaid hospitalizations, emergency department (ED) visits, and readmissions. Combining prior evidence from HCIA R1 with the estimates from the frequentist regressions for the CCRI led to a Bayesian estimate of the program's impact on total Medicare expenditures of 5 percent (an estimated increase of $221 per beneficiary per month) in the first year.

Table D.1. Comparison of frequentist and Bayesian impact estimates for CCRI in the first year after enrollment

		Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Follow-up period	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM) (Medicare)	Quarter 1	-93 (-649, 463)	505 (234, 794)	4%	-1%	7%
	Year 1	174 (-104, 452)	221 (6.7, 442)	3%	4%	5%
Hospital admissions (Medicaid)	Quarter 1	-276 (-381, -170)	112 (11, 211)	5%	-15%	6%
	Year 1	8.7 (-61, 79)	39 (-41, 117)	4%	< 1%	4%
ED visits (Medicaid)	Quarter 1	-241 (-594, 113)	461 (0.19, 895)	4%	-3%	6%
	Year 1	553 (281, 826)	168 (-222, 556)	3%	11%	3%
Readmissions (Medicaid)	Quarter 1	-0.01 (-0.02, 0.01)	0.00 (0.00, 0.01)	4%	-15%	6%
	Year 1	0.00 (-0.04, 0.03)	0.00 (0.00, 0.01)	4%	-4%	4%

Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. Readmissions are measured as the percentage of beneficiaries with a readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

Table D.1 (*continued*)

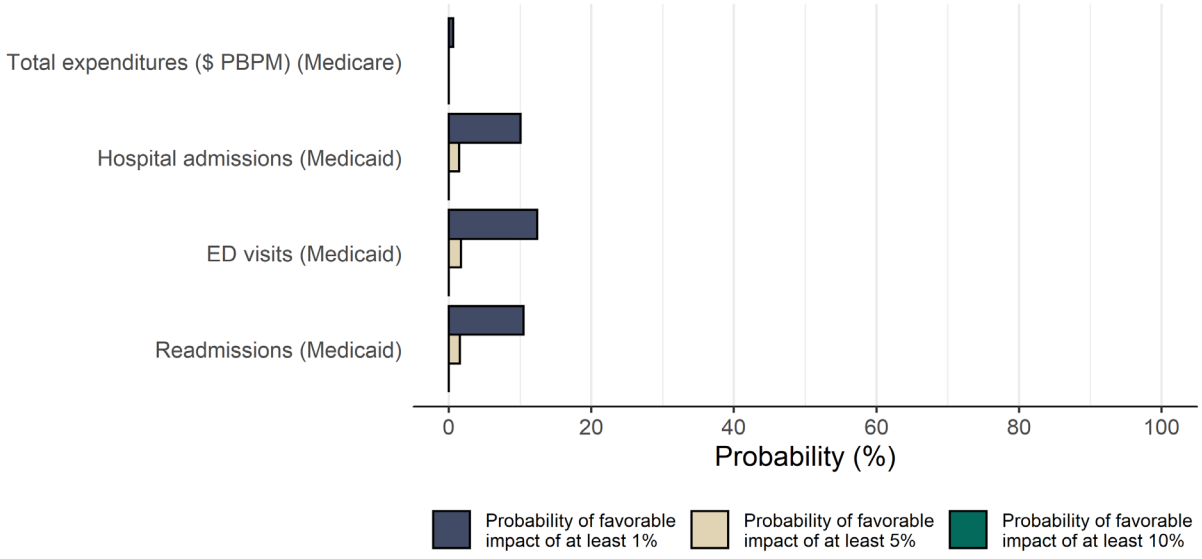
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

CCRI = Community Care Response Initiative; ED = emergency department; HCIA R1= Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results are imprecise due to skewed distributions that increase the variation in the outcomes, the Bayesian model gave more weight to the prior and produced more neutral estimates. By increasing the precision of the estimates, the Bayesian approach has also sharpened some suggestive findings in the frequentist analysis to statistically meaningful findings; for example, in the Bayesian analysis the estimates of the initiative's impact on Medicare expenditures and Medicaid readmissions over the course of the first year differ statistically from zero. At the same time, by shrinking impact estimates toward the range of values observed in HCIA R1, the Bayesian model also attenuated some findings, such as the statistically significant increase in ED visits across the first year; in the Bayesian analysis, the magnitude of the impact is much smaller (168 visits per 1,000 beneficiaries, compared to 553) and the uncertainty interval covers zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that the CCRI achieved favorable impacts during the first year on the four core outcomes at three different thresholds: favorable impacts of 1, 5, and 10 percent or more. Quarterly impact probabilities were negligible and are therefore not presented here.

Figure D.1. Probability that the CCRI program had a favorable impact on key outcomes



Sources: Mathematica's analysis of information from the awardee's program finder file from December 2014 through February 2018 and Medicaid claims and enrollment data from December 2012 through August 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Figure D.1 (*continued*)

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. Readmissions are measured as the percentage of beneficiaries with a readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

CCRI = Community Care Response Initiative; ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a small probability—in the range of 10 percent—that the CCRI had a favorable impact of 1 percent or more on hospital admissions, ED visits, and readmissions in the Medicaid population. The probability of favorable impacts of 1 percent or more on total Medicare expenditures is even smaller, less than 5 percent. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the CCRI did not have favorable impacts on total expenditures, hospital admissions, or ED visits over a full follow-up year.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Clifford W. Beers Guidance Clinic

September 2020

Lisa M. Lines, Grace Anglin, Justin Kirschner, Michaella Morzuch, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

CLIFFORD BEERS CLINIC

Clifford W. Beers Guidance Clinic, Inc., a community-based mental health clinic in Connecticut, received Round 2 of the Health Care Innovation Awards (HCIA R2) funding to implement Wraparound New Haven, a new program serving children enrolled in Medicaid and their families. The program focused on recruiting Medicaid-enrolled children younger than 18 in Greater New Haven who had a chronic physical health condition, had or were at risk of a mental health condition, and had more than one emergency department (ED) visit or any hospitalization in the prior 12 months. Wraparound New Haven enrolled children meeting these criteria and all interested members of their families for 6 to 12 months. During this time, care coordinators helped families identify health and wellness goals and prioritize and implement strategies for addressing those goals. If needed, Wraparound New Haven's behavioral health clinicians provided bridge counseling services to family members of the primary enrollee until a more permanent source of care was secured. Table 1 summarizes the program's key characteristics.

The awardee designed its program to coordinate services for all interested family members and address the interconnected nature of children's behavioral and physical health needs, which family stressors—such as housing or food insecurity—often exacerbated. Clifford Beers expected the program would lead to improvements in coordination of medical and behavioral health care across multiple care settings, increased social connections and supports, and increased family engagement in care. These intermediate outcomes, in turn, would result in improvements in physical and mental health status and reductions in fragmentation of services and the cost of care.

Important issues for understanding the evaluation

- Wraparound New Haven connected eligible children with complex needs and their families to care coordinators to better manage, coordinate, and integrate behavioral and physical health services and social supports
- A rigorous impact evaluation of Wraparound New Haven was not possible because (1) a valid comparison group could not be defined because the program's key eligibility criteria (having a mental health diagnosis or a diagnosis that predicted mental health issues).could not be fully captured in the claims data and (2) the number of participants who did meet eligibility criteria from claims data was too small to yield adequate statistical precision.
- Many of the outcomes that Clifford Beers expected to affect (such as housing security) could not be measured in the Medicaid claims data.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Wraparound New Haven connected eligible children with complex needs and their families to care coordinators to better manage, coordinate, and integrate behavioral and physical health services and social supports.
Major innovation	Wraparound New Haven was innovative in that it focused on the entire family unit's behavioral and physical health needs, rather than a single participant's behavioral health needs.
Program components	• Care management services • Integrated behavioral and physical health care services • Participant and family engagement
Target population	Wraparound New Haven provided services to primary enrollees—children who met the eligibility criteria—and to all members of their families who wished to participate. The primary enrollees had to meet the following criteria:[a] • Resident of Greater New Haven, Connecticut • Younger than 18 • Current Medicaid beneficiary • At least one chronic physical health diagnosis (broadly defined as any condition that consistently affected a child's health status) and one mental health diagnosis or living with conditions that tended to predict mental health conditions • Either two or more ED visits or a hospitalization during the prior 12 months
Participating providers	1 clinic in New Haven, Connecticut
Level of engagement	Wraparound New Haven program staff kept participants actively engaged in the program.
Theory of change or theory of action	Clifford Beers hypothesized that families working closely with a care coordinator would have a better understanding of how to manage their own health, determine the services they needed, and find those services. Accessing these services would, in turn, result in improved mental and physical health outcomes and lower health care spending.
Award amount	$9,739,427
Effective launch date	December 2, 2014
Program settings	Community-based settings and home
Market area	Greater New Haven, Connecticut
Target outcomes	• Improve the coordination of physical and behavioral health care • Enhance family engagement with providers • Improve participants' physical and mental health status • Increase participants' social connections and social supports • Reduce health care spending
Payment model	Clifford Beers reached a fee-for-service reimbursement arrangement with a major private insurer for Wraparound New Haven services. It also pursued a value-based, capitated payment arrangement with the Connecticut State Department of Social Services.

Table 1 (*continued*)

Program characteristics	Description
Sustainability plans	The awardee integrated the key element of the program (in-home care coordination services) into the State Innovation Model and extended the program to a broader group of privately insured children (in terms of geography, age, and condition). Through contracts with other state agencies, the awardee also piloted the program for elderly individuals living with their families, and integrated components of the program into a new center for children and adults with autism.

[a] The awardee, with CMS's approval, changed the eligibility criteria in program Year 1 from requiring the primary enrollee to have three or more visits to the ED to two or more visits and expanded the service area from New Haven proper to Greater New Haven (17 cities and towns).

CMS = Centers for Medicare & Medicaid Services; ED = emergency department.

A rigorous impact evaluation to test whether the program achieved its aims was not possible because a key component of the eligibility criteria (having a mental health diagnosis or a diagnosis that predicted mental health issues) for the treatment group could not be replicated in the Medicaid claims data, and the subset of participants meeting claims-based eligibility criteria was too small to yield estimates of adequate precision. In addition, many of the outcomes that Clifford Beers expected to affect could not be measured using Medicaid claims data. Table 2 summarizes the key features of the descriptive analysis.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	A rigorous impact evaluation to test whether the program achieved its aims was not possible because key components of the eligibility criteria for the treatment group could not be replicated in the Medicaid claims data, and the subset of participants meeting claims-based eligibility criteria was too small to yield estimates of adequate precision. As a result, this report is limited to describing the demographic and health characteristics of Medicaid participants before they enrolled in the program.
Intervention group for descriptive analysis	Although the program's total enrollment was 1,944, the finder file with Medicaid identifiers included only 1549 unique beneficiaries; of these, 1,031 of these were family members and only 518 were primary enrollees. This analysis used the 518 primary enrollees as a starting point. Of the 518 primary enrollees, the analysis excluded 49 because they lacked sufficient baseline Medicaid data and 164 because they did not meet one or more of the program eligibility criteria. Thus, the final sample for this descriptive analysis included 305 primary enrollees.
Limitations	Due to the problems noted above, no inferences can be made about the impact of this program on Medicare costs or other program outcomes.

PROGRAM DESIGN AND ADAPTATION

Wraparound New Haven included three main components: (1) care management services, (2) patient and family engagement, and (3) integrated care services.[1]

Care management services

Care coordinators helped families identify and access services to address their physical and behavioral health and social service needs (such as housing or employment). To support their work, Wraparound New Haven clinicians completed a series of assessments to identify families' needs and strengths. Drawing on these results, care coordinators worked with families to develop and implement a care plan to address their medical, behavioral, and social goals. When needed, care coordinators also provided referrals to clinical and community-based services and supports and, with permission of the family, shared the care plan with relevant providers. Care coordinators attempted to hold meetings—about once per month—with the participants' primary care physician and other community-based providers and organizations or individuals that supported the participants to gain input for the care plan and support for achieving the family's goals. The program served enrolled families for 6 to 12 months.

Integrated care services

The awardee sought to coordinate all physical and behavioral health services for enrolled families. To do so, physical health providers and behavioral health clinicians on staff and the care coordinators coordinated care with providers in the community. Physical health providers on staff worked closely with care coordinators to conduct case reviews and medication reviews for enrollees. The program nurse conducted follow-up calls with families after a family member had an ED visit or inpatient hospitalization to ensure that they scheduled a follow-up appointment and to identify strategies to help the family avoid another ED visit or hospitalization. During the first year of the program, staff behavioral health clinicians started providing short-term counseling services after the care coordinators indicated that finding counseling services for participants was challenging. These clinicians also helped patients secure long-term counseling from other providers as needed. During the third year of the program, the awardee hired a nutritionist after recognizing that participants with hypertension, diabetes, and high blood pressure needed additional support.

Engaging patients and families

Unlike many care coordination programs focused on a single participant, Wraparound New Haven viewed families' needs as interconnected and sought to engage and coordinate services for the entire family unit. Care coordinators attempted to empower families in their navigation of the physical and behavioral health systems, not to manage those services for them. Care

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

coordinators worked closely with families to identify their health and wellness goals, prioritize strategies for addressing those goals, and develop their care plans. Families reviewed their care plans with their care coordinators on a regular basis and made any necessary adjustments.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Care coordinators and awardee leaders reported that the program overcame initial challenges in recruitment, and went on to successfully engage participants, eventually meeting their physical and behavioral-health needs. At the outset, the Wraparound New Haven program faced challenges recruiting participants. Initially, the program relied on referrals from community health providers and social service organizations. To increase enrollment, in October 2015, the awardee embedded a nurse at a major potential referral site (Yale New Haven Hospital Primary Care Center) to help identify potential enrollees and facilitate recruitment. The awardee also relaxed the prior eligibility criteria for the primary enrollee from three to two ED visits and expanded the residency criterion from New Haven proper to Greater New Haven (17 cities and towns). Although the awardee increased enrollment over time, it fell short of its goal of 2,284 participants, instead enrolling 1,944 participants by August 2017 (including both primary enrollees and family members). After they enrolled, the program engaged participants through frequent contacts with their care coordinators. Participants met with their care coordinators at the frequency planned by Clifford Beers, which was about two or three times per month.

By the middle of the second program year, participants had received most services and supports as intended. In Year 1, care coordinators worked with families largely on their behavioral health needs—instead of focusing on both behavioral and physical health needs. Later, Clifford Beers sharpened its focus on physical health services, particularly by hiring part-time medical staff (two pediatricians, an internist, and a nurse) to support the care coordinators. Medical staff educated care coordinators on common chronic physical health conditions, encouraged them to broach physical health topics with families, and conducted medical record and case reviews for participants. Also, during the second program year, the awardee created two

Implications of program implementation for achieving program goals

- One-half of the analytic sample enrolled in Wraparound New Haven before Clifford Beers fully implemented the program and, thus, might not have fully benefited from the program.
- Many families focused on intermediate outcomes—such as losing weight or improving housing security—that were anticipated to result in improved long-term health outcomes. However, these impacts might not be observed during the limited two-year program period covered by the evaluation.
- Nonetheless, program staff felt that Wraparound New Haven had a positive impact on families' understanding of how to manage their physical and behavioral health needs.

physical health tracks, one for participants with complex physical health needs (which included primary enrollees and their family members with complex needs who opted in) and one for those without complex physical health needs. Participants in the comprehensive physical health track were to receive ongoing medical record and case reviews by Wraparound New Haven staff physicians and monthly check-ins from care coordinators and participants' primary care physicians.

Care coordinators and awardee leaders reported that, as a result of these efforts, the program successfully addressed participants' physical and behavioral health needs by the middle of the second program year. As evidence of this change, the rate of contacts between care coordinators and primary care physicians or specialists increased from a median of 0.5 per family per month in the first half of the program to 0.8 per family per month in the second half.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Study sample

Although the awardee enrolled 1,944 participants, the descriptive analysis focused on the 518 primary New Haven participants, excluding the 1,031 family members of the enrollees and 395 who were not included in the finder file of Medicaid beneficiaries for the evaluation. The sample was then limited to those enrollees with sufficient Medicaid data in the baseline period who met the program's eligibility criteria that could be replicated in the Medicaid claims data. Of the 518 primary enrollees, the analysis excluded 49 enrollees because they lacked sufficient baseline Medicaid data and 164 because they did not meet one or more of the program eligibility criteria according to the Medicaid claims and enrollment data. Of these, the analysis excluded 72 enrollees because they did not have at least one chronic physical health diagnosis, broadly defined as any condition that consistently affected a child's health status. It excluded another 23 because they did not have either two or more ED visits or any hospitalization during the prior 12 months. Finally, it excluded 30 enrollees because they did not live in New Haven county and 39 because they were older than 17. After these exclusions, 305 Medicaid participants remained in the descriptive analysis (Appendix A, Table A.1).

Characteristics of Medicaid participants

In line with Clifford Beers' goals for the program, the descriptive analysis indicates that Wraparound New Haven served children with significant health needs and high health care service use during the year before enrolling in the program (Table 3). All enrollees included in the descriptive analysis had (by definition) a chronic physical health condition, most commonly a pulmonary condition (61 percent), cardiovascular disease (28 percent), or a skeletal condition (27 percent). Moreover, 40 percent of enrollees in the analytic sample had at least one Medicaid claim for mental health. The Wraparound New Haven program included participants who had a mental health condition or lived in an environment that would be predictive of mental health

concerns; the 60 percent of enrollees without mental health claims likely fall in the latter category.

Given that enrollees had to have had at least one or more hospitalizations or two or more ED visits to participate, participants had a high rate of acute hospitalizations and ED visits during the year before enrollment. Specifically, the hospitalization rate was 948 per 1,000 beneficiaries and the ED visit rate was 2,949 per 1,000 beneficiaries. In turn, Medicaid expenditures per month ($1,777) were much higher than the average for all Medicaid children in Connecticut ($281 per month).

Table 3. Baseline characteristics of Medicaid participants

Measure	Medicaid participants (N = 305)
Demographic, %	
Age group	
Younger than 5 years	41
Age 5 to 11	39
Age 12 to 17	20
Male	58
CDPS score[a]	
Mean	2.1
25th percentile	1.0
Median	1.5
75th percentile	2.4
CDPS condition, %	
Cardiovascular disease	25
Psychiatric condition	40
Skeletal condition	27
Central nervous system condition	20
Pulmonary condition	61
Gastrointestinal condition	21
Diabetes	2.4
Skin condition	10
Renal condition	7.0
Substance abuse	1.3
Cancer	5.5
Developmental disability	8.3
Genital condition	4.4
Metabolic condition	24
Pregnancy	0.83
Eye condition	1.4
Cerebrovascular condition	2.4
AIDS or other infectious disease	7.5

Table 3 (*continued*)

Measure	Medicaid participants (N = 305)
Hematological condition	5.3
Baseline service use and expenditures during the year before enrollment	
Acute hospitalization admissions (per 1,000 beneficiaries)	948
Outpatient ED visits (per 1,000 beneficiaries)	2,949
Primary care visits (per 1,000 beneficiaries)	3,737
Specialty care visits (per 1,000 beneficiaries)	19,001
Total Medicaid expenditures ($ PBPM)	1,777

Sources: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data from December 2014 through August 2017, and Medicaid claims and enrollment data for December 2013 through February 2018, as of December 2019.

Notes: The analysis defined the baseline year as the 12 months before each beneficiary's enrollment date. It defined the enrollment date as the date on which the participant signed a consent form in the ED. It measured all beneficiaries' characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; PBPM = per beneficiary per month; TANF = Temporary Assistance for Needy Families.

Challenges of estimating program impacts

It was not possible to conduct a rigorous impact evaluation of the Clifford Beers program for several reasons. One of the key program eligibility criteria (having a mental health diagnosis or a diagnosis that predicted mental health issues) could not be fully replicated in the claims data. Only 40 percent of the sample had a Medicaid claim for treatment of a mental health diagnosis in the year before enrollment, so 60 percent of the sample presumably lived in an environment that was conducive to mental health issues, but this could not be confirmed in the Medicaid claims data. Limiting the sample to participants who had a confirmed mental health diagnosis in Medicaid claims and met all of the other observable eligibility criteria left only 122 participants, which was too small a sample for analysis. Finally, conducting an analysis that included all patients who met the other program eligibility criteria (that is, all of the criteria listed above that are not related to mental health) was not possible because only 7 percent of these eligible beneficiaries participated. The low participation rate would have prevented detecting statistically significant impacts of reasonable magnitude.

CONCLUSION

By the middle of the second program year, Clifford Beers had successfully implemented Wraparound New Haven. Although the awardee struggled with recruitment, the program eventually enrolled participants with significant health care needs. Program enrollees met frequently with Wraparound New Haven staff, which provided a range of services, including self-management support, care coordination, medication reviews, and behavioral health

counseling. The study could not conduct a rigorous impact analysis because key program eligibility requirements could not be replicated in the Medicaid claims data, and the number of participants whose eligibility could be confirmed in the claims data was too small to yield adequate power to detect even very large true impacts.

PROGRAM SUSTAINABILITY

When its award ended in February 2018, Clifford Beers could not sustain the Wraparound New Haven program for the original target population because of a lack of funding. However, the awardee integrated in-home care coordination services (a key element of the program) into the State Innovation Model and, with foundation funding, used its cost analysis results to inform and develop a strategic plan to pursue state funding.

Clifford Beers scaled up the program to new populations by the end of its award. The awardee extended the program to a broader group of privately insured children (in terms of geography, age, and condition). Through contracts with other state agencies, the awardee also piloted the program for elderly individuals living with their families and integrated components of the program into a new center for children and adults with autism spectrum disorder and other developmental disabilities.

Clifford Beers Guidance Clinic's proposed payment models

The awardee originally designed two payment arrangements to sustain Wraparound New Haven for children with Medicaid: (1) a value-based arrangement with the state Medicaid agency, potentially including shared savings; and 2) contracting with provider organizations participating in value-based arrangements, for which care coordination might help reduce costs. The awardee later developed a third payment model consisting of a fee-for-service payment for implementing the program to privately insured children.

The awardee's attempts to implement its original payment model for Medicaid beneficiaries were unsuccessful. The awardee attributed the lack of progress on its Medicaid payment model primarily to state budget shortfalls and lack of data on program outcomes. However, the awardee established a contract with the state's largest commercial insurer, Anthem Blue Cross Blue Shield, to implement the model for commercially insured children and their families, under the name Advanced Care Coordination, or ACCORD. At the end of the award, Clifford Beers pursued further contracts with commercial insurers and expected that receiving accreditation from The Joint Commission for the program would facilitate these efforts.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Although 1,944 participants enrolled in the program, the finder file included only 1,549 distinct individuals (518 primary enrollees and 1,030 family members) with Medicaid identifiers provided for the evaluation. The starting point for this analysis were the 518 primary enrollees in this finder file, excluding the 1,031 family members from the analysis. Next, the analysis was limited to the 354 of those primary enrollees who met the awardees' eligibility criteria that could be assessed with Medicaid claims and enrollment data. This requirement led to excluding 23 without enough inpatient stays or emergency department visits during the 12 months before enrollment, 39 who could not be confirmed as 17 or younger, 30 who did not live in New Haven County, and 72 with no chronic conditions listed in claims during the baseline year. Clifford Beers limited participation in the program to Medicaid enrollees who had a mental health condition or lived in an environment that would be predictive of mental health concerns. Medicaid data do not provide information on individuals' living environments, so the analysis could not use this criterion to identify the analytic sample. Finally, the analysis excluded 49 participants from the sample due to limited Medicaid claims data availability over the study period. The final analysis sample includes the 305 primary enrollees who met the eligibility criteria who could be confirmed in the Medicaid claims data and had sufficient claims data for analysis.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total primary enrollees and family members		**1,944**
Duplicated primary enrollees and family members	395	1,549
Excluded family members from analysis	1,031	518
Excluded enrollees because edibility criteria could not be replicated:[a]		
Did not have inpatient stay or at least 2 ED visits or observation stays during 12 months before enrollment	23	495
Older than 17 at the enrollment date or missing age data	39	456
Not a resident of New Haven County in the month of enrollment	30	426
No chronic conditions during the year before enrollment	72	354
Did not have any Medicaid claims during the month of enrollment	10	344
Not continuously enrolled in Medicaid FFS during 3 months before enrollment	39	305
Final analytic sample		**305**

Sources: Mathematica's analysis of information from the awardee's program finder file and Medicaid claims and enrollment data from January 1, 2012, through August 31, 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

[a] Clifford Beers included participants in its program who had a mental health condition or lived in an environment that would be predictive of mental health concerns. The analysis could not use Medicaid claims and enrollment data to determine if an individual lived in an environment that would be predictive of mental health concerns, so it did not use this aspect of the eligibility criteria to identify the analytic sample.

Table A.1 (*continued*)

[b] The 18 chronic condition categories included AIDS or other infectious disease, cancer, cerebrovascular condition, central nervous system condition, cardiovascular disease, diabetes, developmental disability, eye condition, gastrointestinal condition, genital condition, hematological condition, metabolic condition, psychiatric condition, pulmonary condition, renal condition, skeletal condition, skin condition, and substance abuse.

ED = emergency department; FFS = fee-for-service.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Community Care of North Carolina

September 2020

Eva Chang, Molly Lynch, Kevin Smith, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

COMMUNITY CARE OF NORTH CAROLINA

Community Care of North Carolina (CCNC) received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the North Carolina Community Pharmacy Enhanced Services Network (CPESN). This community-based care delivery and payment model focused on transforming community pharmacies by motivating pharmacists to focus on improving medication management for the most at-risk patients, in addition to the pharmacies' traditional role of dispensing medications. The target population included Medicaid. Medicare, and S-CHIP patients attributed to 253 participating pharmacies with one or more chronic medical conditions treated through medication. The program launched in March 2015, seven months after award. The intervention period covered under HCIA R2 ended in May 2018, after receiving a one-year no-cost extension. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program aimed to reduce total health care expenditures and improve chronic disease self-management, adherence to prescribed medication, multidisciplinary care across providers, and hospital follow-up with enhanced pharmacy services for Medicare, Medicaid, and S-CHIP patients in North Carolina with chronic conditions through pharmacy-based medication management.
- CCNC created the CPESN, which aimed to transform community pharmacies in North Carolina by motivating pharmacists to focus on improving medication management for the most at-risk patients.
- This impact analysis relied on 110,968 Medicare fee-for-service (FFS) beneficiaries enrolled in a Part D plan and attributed to one of 253 CCNC treatment pharmacies using claims-based attribution, and 147,034 comparison group beneficiaries attributed to 478 North Carolina pharmacies that had similar characteristics as participating pharmacies but did not participate in the CPESN.

Awardee leaders hypothesized that by shifting pharmacists' incentives from the traditional encounter-based payment model to a value-based model, pharmacists would become more involved in delivering care as they provided enhanced pharmacy services, including medication management, that was commensurate with the needs and health status of their attributed patients. Participating pharmacies used PHARMACeHOME, a pharmacy information exchange platform that enabled pharmacists to understand a patient's prescription history to deliver effective medication management and to support coordination with the extended care team of the patient-centered medical home. The program sought to improve clinical outcomes through chronic disease self-management, adherence to prescribed medication, multidisciplinary care across providers, and hospital follow-up with enhanced pharmacy services, and thereby reduce total annual health care expenditures by at least $30 million by 2017.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The North Carolina Community Pharmacy Enhanced Services Network (CPESN) sought to incorporate integrated medication management strategies into the interactions between patients and community pharmacists while offering pharmacists an incentive to provide enhanced services, collaborate with a patient's care team, and address gaps in care.
Major innovation	The CPESN changed the practice of community pharmacy by incentivizing pharmacists to move from primarily filling and dispensing medications to providing enhanced services that address gaps in patients' care by improving medication management for the most at-risk patients.
Program components	• Enhanced medication therapy management services delivered in community pharmacy settings • Comprehensive initial pharmacy assessments (CIPAs) and identification of drug therapy problems (DTPs) • Integrated care management and medication management information platform to enhance care delivery by involving pharmacists more directly in patients' care • Payment model that provided pharmacies with incentives to provide care to high-risk patients and address gaps in care
Target population	The awardee sought to engage Medicaid enrollees or Medicare beneficiaries who had one or more chronic medical conditions being treated through medication or who were identified by either a referring physician or the pharmacy itself as likely to benefit from the medication management intervention.
Total enrollment	The awardee passively enrolled 388,053 Medicare, Medicaid, and S-CHIP beneficiaries from March 2015 through May 2018, representing 109 percent of its original enrollment goal.
Theory of change or theory of action	The awardee hypothesized that enhancing the role of pharmacists in the ongoing management of medication for patients with chronic disease would improve the quality and coordination of care and, in turn, reduce avoidable inpatient and ED service use and lower the total cost of care.
Award amount	$15,634,150
Effective launch date	The program began in March 2015 when it started attributing patients to participating pharmacies.
Program settings	Participating community pharmacies
Market area	North Carolina
Target outcomes	• Increase self-management of chronic diseases • Reduce the number of patients who do not comply with prescribed medications by 28 percent • Increase multidisciplinary care across providers—for example, by increasing the proportion of patients with a completed referral and follow-up care • Increase the rate of hospitalized patients who receive enhanced pharmacy services within 72 hours of discharge • Reduce total annual health expenditures among participants by at least $30 million
Payment model	New risk- and value-based payments combined with PBPM payments for care management and coordination services
Sustainability plans	CCNC is actively engaging the state Medicaid program and other payers to fund the program after this award ends. Funding is a challenge to sustainability until CCNC finalizes agreements with payers.

CCNC = Community Care of North Carolina; CPESN = Community Pharmacy Enhanced Services Network; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; S-CHIP = State Children's Health Insurance Program.

This impact analysis presented in this report includes 110,968 Medicare FFS beneficiaries who were enrolled in a Medicare Part D plan, attributed to one of the 253 CCNC treatment pharmacies, and met standard claims-based study inclusion criteria. The comparison group included 147,034 Medicare FFS and Part D beneficiaries attributed to 478 matched North Carolina pharmacies that did not participate in the intervention. Table 2 summarizes the key features of the impact evaluation. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The impact estimates rely on a difference-in-differences study design, where the intervention group includes all eligible Medicare patients based on Medicare Part D data. This design measures program effects as the change in outcomes among the intervention group before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Enrollment in the study is assumed to be in the month when the patient is first attributed to a treatment or comparison pharmacy after program started up on March 1, 2015.
Intervention group for evaluation	The intervention group for the evaluation relied on 110,968 Medicare FFS beneficiaries enrolled in a Medicare Part D plan and identified through claims-based attribution to a CCNC treatment pharmacy from March 2015 through May 2017. About 55 percent of the claims-based attribution study sample appeared on CCNC's file of intervention participants. Additional criteria for inclusion in the claims-based attribution sample included meeting the standard Medicare enrollment-based inclusion criteria (enrolled in both Medicare Parts A and B, not enrolled in Medicare Advantage, and Medicare was the primary payer) and meeting the CCNC attribution inclusion criteria (filled a chronic disease medication from a treatment pharmacy and had 80 percent or more of their prescription fills from a given pharmacy in the three months before attribution).
Comparison group	The comparison group included 147,034 Medicare FFS beneficiaries enrolled in a Medicare Part D plan who were identified through Medicare claims-based attribution to one of the 478 comparison pharmacies from March 2015 through May 2017, and who were eligible for Medicare for at least 90 days prior to the attribution date and did not die within 30 days of attribution. Comparison pharmacies were selected by identifying pharmacies that closely resembled the participating pharmacies. Propensity score matching occurred at the beneficiary level.
Limitations	The evaluation estimated program impacts using the eligible population identified through Medicare claims-based attribution rather than the just the treatment participants identified by CCNC. Because only 55 percent of the Medicare claims-based attribution beneficiaries appear on CCNC's file of program participants, the estimated program effects on eligibles may be only about half as large as the effects on participants. Impact estimates may be biased if outcomes for comparison beneficiaries change at different rates from outcomes for treatment beneficiaries due to differential changes in other factors that affect the two groups of beneficiaries, such as local market factors.

CCNC = Community Care of North Carolina; FFS = fee-for-services.

PROGRAM DESIGN AND ADAPTATION

To achieve CCNC's aim to expand the role of pharmacists in the ongoing management of patients with chronic diseases, the program's service delivery model focused on two key components: (1) enhanced medication therapy management services in community pharmacy settings and (2) integrated care management and a medication management information

platform. The program included a payment model that provided pharmacies with incentives to provide care to high-risk patients and address gaps in care.[1]

Delivering enhanced pharmacy services in community pharmacy settings

Pharmacists in the CPESN provided a broad spectrum of enhanced services to facilitate pharmacy care management. Community pharmacy care management (CPCM) was the most intensive service delivered by the CPESN pharmacists and deemed essential for achieving CCNC's target outcomes. The CPCM included a comprehensive initial pharmacy assessment (CIPA), which was both a clinical review of the patient's medical conditions and drug therapies and a review of the barriers to optimal care—such as low health literacy, environmental factors like lack of transportation, and lack of caregiver or in-home supports—to enrolled patients who were most at-risk for medication-related complications. Critical to CPCM was the patient receiving regular follow-up medication management consultations at intervals deemed necessary by the pharmacist, who updated the care plan and coordinated follow-up activities with other care team members. In addition to the CIPA, pharmacists also identified patients with drug therapy problems (DTPs), such as drug-drug interactions or allergies. Although less intensive than CIPA, pharmacists reached out to patients with DTPs for a conversation and potential follow-up with providers.

All participants could have received a variety of other enhanced pharmacy services. The additional types and levels of enhanced pharmacy services varied considerably based on pharmacies' resources and participants' needs. Enhanced services included medication reconciliation, home visits, medication delivery, adherence packaging, referrals for behavioral or mental health services and for home and community-based services, and support for improving a participant's self-management and informed decision making. Notably, the types and levels of services each pharmacy offered also varied before their involvement in the CPESN.

Integrated care management and medication management information platform

During the first two years of the program, CCNC developed and implemented PHARMACeHOME, a pharmacy information exchange platform. This system aimed to allow pharmacists to use a patient's prescription history to deliver effective medication management services. Pharmacists used the system to document activities involved in providing enhanced services and submit this information to the awardee for monthly payment. They also used it to support the coordination of care by serving as an extension of a care manager. Late in the third year of the program (starting in June 2017), CCNC transitioned the CPESN pharmacies to eCare Plans, a shared, real-time electronic platform that documented the process pharmacists use to assess and provide care to a patient. Pharmacists considered eCare Plans to be an improvement because the health information technology (IT) platform was integrated into pharmacists' workflows and easier to use than PHARMACeHOME.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Payment model to incentivize provision of enhanced pharmacy services

CCNC provided incentives to pharmacies to provide CIPAs and identify participants with DTPs through a payment model. During the first two years of the program, pharmacies could be reimbursed for up to 3 CIPAs per person per year. Pharmacies also received a per member per month (PMPM) payment for DTP patients. In the third program year, CCNC changed its payment model to encourage pharmacies to spread CIPAs across more patients. The awardee allowed only one CIPA per person per year and required pharmacies to provide at least three CIPAs in total to remain in the CPESN. The awardee also changed DTP payments in the third year by requiring patients to have received a CIPA to be eligible for a DTP payment.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

To successfully implement the service delivery components of the model, the CPESN pharmacies had to augment or change their workflows to accommodate new processes. Although CCNC provided leadership, support, training, and technical assistance to implement the program, the awardee faced challenges with service delivery among participating pharmacies. While interest among participating pharmacies in this new service delivery model was high, the intervention required pharmacies to make major workflow changes; adapt to an ever-changing payment model; and use a burdensome health IT system, PHARMACeHOME, to manage care. As a result, CCNC noted substantial differences across pharmacies in how they delivered services.

During the first program year, CCNC realized that the CPESN pharmacies entered the network in various stages of readiness to make the workflow changes necessary to support the intervention. Given the range of pharmacies' readiness, the awardee provided them with one-on-one support from a CCNC pharmacy technician and pharmacist through a helpline and conducted site visits to help struggling pharmacies in the first year of the program. Program data reflected this ramp-up in the effectiveness of service delivery, showing that among the almost 17,000 CIPAs provided, about 3,300 were provided in Year 1 compared to 8,700 in Year 2, and 7,700 in Year 3. Despite the intensive assistance provided by the awardee, several pharmacies continued to have low levels of engagement in the CPESN, which the awardee measured as having completed few, if any, CIPAs. Because pharmacies also varied widely in how they delivered the intervention services, the awardee found fidelity and pharmacy engagement was difficult to measure and categorize.

Although the awardee constantly collected feedback from participating pharmacies to better implement the intervention, pharmacies reported that every change in the structure of the incentive payments disrupted their workflows. During the three-year intervention, the awardee introduced five iterations of the payment model. From March 2015 to May 2017, the payment model combined PMPM and encounter-based payments. In June 2017, concurrent with the roll-out of eCare Plans, CCNC shifted to a risk- and performance-based payment model. Under this

model, a combination of the patient's final risk score in a given month and the pharmacy's most recent performance score determined pharmacy payments. The multitude of changes in incentive payments was challenging for pharmacies to follow and led to some pharmacies becoming less engaged and offering fewer services over time. These payment model changes also were tied to the use of the health IT platforms. The PHARMACeHOME platform often stymied service delivery; pharmacists considered the system to be complex and time-intensive, which discouraged them from entering data to submit monthly payment information. As CCNC moved toward eCare Plans and away from CIPAs during Year 3, some pharmacies reported they did not invest as much time in providing CIPAs during this transition period because they were waiting for eCare Plans and the concurrent final alternative payment model to be finalized.

Implications of program implementation for detecting impacts

- Because the CPESN pharmacies entered the network in various stages of readiness, CCNC had to provide intensive training and other assistance to pharmacies to help them modify their workflows to support intervention services.
- Pharmacies varied widely in how they delivered services; fidelity and pharmacy engagement were difficult to measure and categorize.
- Multiple changes to the payment model over time created different incentives that affected how pharmacies delivered services.
- Pharmacists found the PHARMACeHOME platform (used to document service delivery and submit payment information) complex and time consuming.

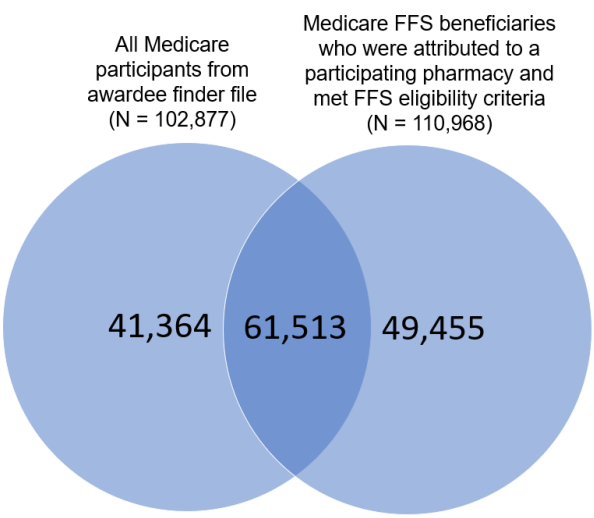
ESTIMATING PROGRAM IMPACTS

Pharmacy selection

Community pharmacies enrolled into the CPESN in the first program year through a CCNC statewide call to all community-based pharmacies; subsequent identification of pharmacies interested in the CPESN occurred through word-of-mouth. For estimating program impacts, treatment pharmacies were CPESN pharmacies that participated in the payment model for at least one program year. During the three program years, the study identified 253 treatment pharmacies. Comparison pharmacies were selected from a list of more than 3,000 active nonparticipating pharmacies obtained from the North Carolina Board of Pharmacies. The study excluded pharmacies if they had characteristics not found among the intervention pharmacies (for example, they did not have a full permit or were a nuclear pharmacy). The evaluation then selected North Carolina-based comparison pharmacies that matched treatment group pharmacies on pharmacy type (independent, chain, free clinic, or other) and National Provider Identifier primary taxonomy (such as retail pharmacy, clinic pharmacy, specialty pharmacy) as well as geographic proximity to an intervention pharmacy and a selected few other pharmacy-specific variables. The process yielded 478 comparison pharmacies.

Beneficiary attribution and study sample

CCNC patients were enrolled monthly into the program through two methods. The first method relied on claims-based attribution from pharmacy fill information provided by the intervention pharmacies. The awardee attributed the patient to a treatment pharmacy because the patient had one or more chronic condition medications filled by that pharmacy and 80 percent or more of those medications filled at that pharmacy in the 90 days before the attribution date. In the second method, a pharmacist relied on clinical judgment to identify and document that a patient had a DTP. CCNC attributed beneficiaries monthly, enabling a participant's treatment status and pharmacy to change throughout the program. To estimate program impacts, the study attributed beneficiaries to a treatment or comparison pharmacy by replicating the CCNC attribution algorithm but using Medicare claims data, including Part D data. However, the evaluation kept beneficiary treatment status and treatment pharmacy constant. When the evaluation attributed a beneficiary to a treatment pharmacy, it deemed the beneficiary a treatment beneficiary for the remainder of the study. Similarly, beneficiaries attributed to a comparison pharmacy remained in the comparison group for the duration of the follow-up, with one exception. If the analysis first attributed a beneficiary to a comparison pharmacy and later to a treatment pharmacy, it considered the beneficiary a treatment beneficiary and removed him or her from the comparison group to avoid contamination.



The study estimated program impacts over 110,968 Medicare FFS beneficiaries who were eligible to participate in the intervention as determined through Medicare claims-based attribution, rather than the 102,877 Medicare participants identified by the awardee using information not available in claims data to select a comparison group: the use of Medicaid claims data and pharmacy fill data provided by the treatment pharmacies; and the use of clinical judgment by pharmacists and providers to identify beneficiaries at-risk for medication-related complications. For this reason, an intent-to-treat approach was used to reduce the potential for biased impact estimates due to possible systematic differences in the treatment and comparison beneficiaries and improves the generalizability of the findings. Of the Medicare claims-based attributed beneficiaries included in the study sample, 55 percent (61,513) participated in the intervention. The relatively high participation rate among all eligible beneficiaries and the large sample size ensures a high likelihood of detecting impacts of modest size.

The impact analysis relied on 110,968 Medicare FFS beneficiaries identified through Medicare Part D claims-based attribution to a CCNC treatment pharmacy from March 1, 2015, through

May 31, 2017. The study sample excluded Medicaid-only beneficiaries, State Children's Health Insurance Program (S-CHIP) beneficiaries, and Medicare FFS beneficiaries who were not enrolled in a Medicare Part D plan. The program also sought to enroll high-risk individuals identified by pharmacists. Pharmacists identified these high-risk individuals as having a DTP or in need of a CIPA and they comprised about 10 percent of all participants. Although the claims-based attribution might have included some of these individuals, Medicare claims could not replicate the identification of high-risk comparison beneficiaries; thus, all claims-attributed beneficiaries were included in the treatment and comparison groups. Finally, in the final program year, CCNC moved toward eCare Plans and away from using CIPAs and DTPs as an eligibility criterion. Because eCare Plans represented a significant change in service delivery and payment and were implemented near the end of the award period, the impact analysis did not include patients attributed after eCare Plans were introduced in June 2017.

Of the 111,453 Medicare FFS beneficiaries who met the claims-based attribution criteria for inclusion in the evaluation treatment group, the study excluded 317 beneficiaries who were not eligible for Medicare in the 90 days before or on the attribution date and 168 beneficiaries who died within 30 days of attribution. (Appendix A, Table A.1 describes the final sample selection process for the treatment group.) The evaluation compared outcomes for the treatment group to those of the 147,034 comparison Medicare FFS beneficiaries attributed to a comparison pharmacy and propensity score matched to treatment group beneficiaries.

Characteristics of treatment and comparison group beneficiaries

Comparing treatment and comparison beneficiary characteristics at baseline confirmed that the two groups were similar across most baseline characteristics (Table 3), with a few exceptions. The average age of treatment and comparison group members at the time of attribution was 68 and 70, respectively, and treatment beneficiaries were less likely to be age 85 or older (10 vs 14 percent). Treatment beneficiaries were slightly less likely to be White and to be dually eligible for Medicare and Medicaid and more likely to have obtained Medicare eligibility through disability. The average hierarchical condition category (HCC) risk score was 1.3 for the treatment group and 1.4 for the comparison group, indicating expected health care expenditures for both groups were 30 to 40 percent higher than the national average for the full Medicare population. Appendix B provides the full comparison of key baseline characteristics for the treatment and comparison groups measured during the 12 months before attribution.

Most baseline service use measures were similar between the two groups, but there were a few notable differences. During the baseline year, about 20 percent of both groups had an acute care hospitalization and 35 to 40 percent had an outpatient emergency department (ED) visit. Baseline rates of specialty care visits were also similar between treatment and comparison groups. However, both average number of hospitalizations and average Medicare spending per beneficiary per month (PBPM) were about 15 percent greater for the comparison group than for the treatment group. Primary care visits were also higher for the comparison group, by 11 percent. The somewhat higher service use and costs of the comparison group may reflect the

comparison group's higher proportion of patients who are age 85 or older, slightly greater average HCC score, and/or location in areas with different practice patterns.

Table 3. Baseline characteristics treatment and comparison group beneficiaries

Characteristic	Treatment (N = 110,968)	Comparison (N = 147,034)
Demographics		
Age at attribution, years	68	70
Age group, %		
Younger than 65 years	30	27
65 to 74 years	37	34
75 to 84 years	23	24
85 years and older	10	14
Female, %	59	60
White, %	73	76
Original reason for Medicare eligibility, %		
Old age and survivor's insurance	57	59
Disability insurance benefits	42	40
End-stage renal disease	1	1
Medicare/Medicaid dual status	41	45
HCC score[a]		
Mean	1.3	1.4
25th percentile	0.6	0.64
Median	0.96	1.0
75th percentile	1.6	1.8
Service use and expenditures during the year before enrollment		
Any ED visits, %	37	38
Any hospitalizations, %	20	23
Hospital stays (per 1,000 beneficiaries)	337	383
ED visits (per 1,000 beneficiaries)	903	895
Primary care visits, and setting, (per 1,000 beneficiaries)	7,346	8,194
Specialty care visits, any setting, (per 1,000 beneficiaries)	8,568	8,696
Total Medicare expenditures ($ PBPM)	887	1,024

Sources: Mathematica's analysis of CCNC-provided data on pharmacies and information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019.

Notes: The evaluation defined the baseline year as the 365 days before each beneficiary's attribution date. It defined the attribution date as the first month a beneficiary is attributed to a treatment or comparison pharmacy. It also measured all beneficiary characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

All differences except rates of ED visits are statistically significant at the $p < 0.01$ level. Given the large numbers of beneficiaries in each of the study samples, small differences are statistically significant but have little clinical or evaluation importance.

Appendix B presents the full comparison results. ED visit measures include observation stays.

Table 3 (*continued*)

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact estimates depend on a difference-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment (defined for each individual as of their attribution date[2]) relative to the change in outcomes among a comparison group with similar characteristics over periods defined in the same manner. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. However, even if the two groups differ on some unobservable characteristics, to the extent that these characteristics are constant over time, the difference-in-difference model will "net out" the effect of these factors on outcomes. The primary outcomes are total Medicare spending, number of hospital admissions, and number of ED visits. Secondary outcomes include number of primary care and specialty care visits. Appendix A describes the statistical models used to estimate the effects of the program, and it identifies the final analytic sample.

IMPACT RESULTS

The estimated effects of the enhanced pharmacy services intervention on Medicare spending and the use of hospital inpatient and ED services were very small (less than 1 percent of the mean) and not statistically significant over the 24-month follow-up period (Table 4). The program did generate a modest but statistically significant estimated decrease in primary care visits, in any setting, for all treatment beneficiaries relative to comparison beneficiaries during this time period. The estimated rate of primary care visits in any setting decreased by 3.6 percent (or 300 visits per 1,000 beneficiaries) more for the treatment than the comparison group. The full results of the impact analyses and sensitivity analyses are presented in Appendix C. The direction and interpretation of the impact results were not sensitive to whether a beneficiary enrolled early or later during the program period (see Appendix C). Results from a Bayesian analysis are shown in Appendix D and support the reported findings.

[2] The evaluation defined the attribution date as the date a beneficiary was first attributed to a treatment or comparison pharmacy. It defined the pre-attribution (baseline) period as the year before each beneficiary's attribution date and the post-attribution period as the two years after.

Table 4. Estimated impact of enhanced pharmacy services on selected outcomes, two-year follow-up period

	Full group Months 1–24
Total expenditures ($ PBPM)	
Impact ($)	10
Percentage impact	< 1%
p-value	0.19
Hospital stays, per 1,000 beneficiaries	
Impact (rate)	0.51
Percentage impact	< 1%
p-value	0.89
ED visits, per 1,000 beneficiaries	
Impact (rate)	-5.1
Percentage impact	< 1%
p-value	0.48
Primary care visits, any setting, per 1,000 beneficiaries	
Impact (rate)	-300***
Percentage impact	-3.6%
p-value	< 0.01
Specialty care visits, any setting, per 1,000 beneficiaries	
Impact (rate)	-33
Percentage impact	< 1%
p-value	< 0.1
Sample size	
Treatment	110,968
Comparison	147,034

Source: Mathematica's analysis of CCNC-provided data on pharmacies and information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019.

Note: Impact estimates are based on the regression-adjusted difference between the randomized treatment and control group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents the full impact estimates. Appendix D presents the results of the Bayesian analysis. ED visit measures include observation stays.

*Significantly different from zero at the 0.10 level, two-tailed test.

**Significantly different from zero at the 0.05 level, two-tailed test.

***Significantly different from zero at the 0.01 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month.

Several factors might explain why the CCNC pharmacy intervention failed to achieve a discernible reduction in overall total expenditures (a target outcome for the awardee), inpatient hospital stays, or ED visits. First, CCNC constantly modified its service delivery and payment model to support the intervention's goals. Although CCNC created and provided trainings, support, and technical assistance to pharmacies to assist in implementing its service delivery model and to develop lessons learned from high-performing pharmacies, many pharmacies were unable (or unwilling) to modify and adapt their workflow processes to accommodate the new

care management and enhanced pharmacy services, and the multiple changes to the payment model. Second, pharmacy staff reported that a key barrier to effective service delivery had been a sudden increase of high-risk patients attributed or referred to high-performing pharmacies.[3] The pharmacies reported that the high-risk patients strained pharmacy resources and affected payment based on performance ratings, making it less likely that the pharmacies would achieve reductions in total expenditures, inpatient hospital stays, and ED visits during the follow-up period.

Main findings from impact evaluation

- The CCNC pharmacy intervention did not result in discernible effects on total Medicare expenditures, inpatient hospital stays, or ED visits over a 24-month follow-up period.
- CCNC appeared to reduce participants' use of primary care visits slightly, suggesting that the intervention's enhanced pharmacy services may have enabled patients to avoid some primary care visits.

The evaluation found modest reductions in primary care visits. These impact estimates suggest that the medication management and other enhanced pharmacy services provided by the intervention might have eliminated the need for some primary care visits. The decrease in primary care visits, however, was not enough to lower total spending.

CONCLUSION

There is no evidence that the medication management and enhanced pharmacy services intervention affected the targeted outcome of total Medicare spending. Nor did it have an impact on the use of hospital inpatient or ED services. Although the intervention did reduce primary care visits by less than 4 percent, the impact was not sufficient to reduce total expenditures. Challenges integrating the intervention into existing workflows, the use of a burdensome health IT system, and uncertainty in the structure of the payment incentives likely limited the effectiveness of intervention services and hindered the awardee's ability to achieve its goal of reducing total spending by $30 million. Furthermore, few study participants received the case management portion of the intervention that the program considered critical for achieving its goals—only 7 percent of the study beneficiaries received a CIPA and only 9 percent received a CIPA or DTP.

Limitations of evaluation

The analysis has several limitations. First, the evaluation calculated program impacts using the population attributed to participating pharmacies, not the treatment participants identified by the awardee (including CCNC-identified attribution, CIPA, and DTP patients). This likely attenuated program effects somewhat, because only 55 percent of the Medicare claims-based attribution beneficiaries participated in the program. However, estimated effects on expenditures, hospital

[3] The awardee identified high-performing pharmacies in site visits and through other technical assistance activities. These pharmacies tended to have similar attributes, including workflows adjusted to provide enhanced services, good communication with physicians and care teams, and regular referrals.

admissions and ER visits were all so small that any meaningful true impacts would not be masked by this intent-to-treat approach. Second, although the use of an intent-to-treat model to estimate impacts reduced the risk of selection bias, the approach might not have eliminated biases associated with unobservable differences in beneficiary, facility, and market characteristics between treatment and comparison pharmacies that are associated with different changes over time in patient outcomes. For example, the average HCC score was 10 percent lower among the treatment group than the comparison group. Similarly, the rate of primary care visits in any setting was 11 percent lower. Third, the impact analysis did not include Medicaid and S-CHIP patients and patients not enrolled in a Medicare Part D plan. Based on the awardee's data, these patients likely comprised about half of the patients who received enhanced services at treatment pharmacies, and their exclusion limits the generalizability of the evaluation. It is possible that the program had discernible effects for these other populations. Finally, the evaluation lacked data on the targeted intermediate outcomes, such as self-management of chronic diseases, adherence to prescription drug treatment, and proportion of patients with enhanced pharmacy services within 72 hours after discharge.

PROGRAM SUSTAINABILITY

After its award ended in August 2018, CCNC no longer provides payments to pharmacies but does continue to work with interested pharmacies on the importance of proper documentation and setting and reevaluating patient-centered goals. Interested pharmacies continue to incorporate aspects of the program even in the absence of payment because they have sought an opportunity to change the way they practice community pharmacology.

CCNC had not secured agreements with payers by the end of the cooperative agreement, but continued pursuing funding from payers after the award. As a result, participating pharmacies experienced a gap in funding, since CCNC ended its payments to pharmacies using award funding in May 2018. Because of the awardee's focus on Medicaid enrollees, the awardee opted to focus obtaining new funding from the North Carolina Division of Medical Assistance by incorporating the program into a Medicaid reform effort slated to take place in 2019.

CCNC proposed payment model

CCNC proposed a value-based payment model to pay for CPESN. Under the model, pharmacies would receive payments from payers for submitting one care plan per patient per month. The monthly payments would range from $2.50 to $40 per care plan, depending on the patient's risk score and the pharmacy's performance on risk-adjusted metrics. During the award period, both measures were calculated by the awardee or its contractor using Medicaid and Medicare claims data. Pharmacies could earn up to 11 points based on the following risk-adjusted performance metrics, which are updated quarterly:

- Total cost of care (3 possible points)
- ED utilization (2 possible points)
- Inpatient hospital utilization (2 possible points)
- Medication adherence (4 possible points across 4 measures)

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays rely on a difference-in-differences approach with beneficiary fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or emergency department (ED) visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or any ED visit during the baseline period. The intervention years are beneficiary specific and defined relative to each beneficiary's date of attribution. Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of outcomes.

Community Care of North Carolina (CCNC) identified about 55 percent of the claims-based attribution study sample as intervention participants using prescription fill data from participating pharmacists. For the evaluation, the attribution sample used Medicare Part D data to attribute Medicare fee-for-service beneficiaries to a CCNC treatment pharmacy from March 1, 2015, through May 31, 2017 (Table A.1). Additional criteria for inclusion in the attribution sample included meeting the standard Medicare enrollment-based inclusion criteria (enrolled in both Medicare Parts A and B, not enrolled in Medicare Advantage, and Medicare was the primary payer) and meeting the attribution inclusion criteria (filled a chronic disease medication from a treatment pharmacy, and had 80 percent or more of their prescription fills from a given pharmacy in the three months before attribution). A small number of beneficiaries (less than 1 percent) were excluded if they did not have 90 days of eligibility before or on the attribution date or died within 30 days of their attribution date.

Table A.1. Identification of final sample for impact analysis for CCNC

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Attributed sample of Medicare FFS beneficiaries[a]		**111,453**
Did not have 90 days of Medicare Part A and B eligibility before attribution date	312	111,141
Was not eligible for Medicare Part A and B on attribution date	5	111,136
Died within 30 days of attribution	168	110,968
Final analytic sample		**110,968**

Source: Mathematica's analysis of CCNC-provided data on pharmacies and information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019.

[a] The attribution sample included Medicare beneficiaries who met the inclusion criteria (enrolled in both Medicare Part A and B, not enrolled in Medicare Advantage, and Medicare was the primary payer) and met the attribution inclusion criteria (filled a chronic medication from a treatment pharmacy, and had 80 percent or more of their prescription fills from a given pharmacy in the prior three months).

CCNC = Community Care of North Carolina; FFS = fee-for-service.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the baseline variables used compare the treatment and comparison groups. The table displays the means of baseline characteristics for the 110,968 treatment beneficiaries and the 147,034 comparison beneficiaries used in the impact analysis. The variables used for comparing the two groups include demographic characteristics (age, gender, and race); Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category [HCC] score and chronic condition indicators; Medicare expenditures in total and by type of service; service use; and pharmacy characteristics. The variables are measured over various specified intervals within the 12 months before attribution in the intervention. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable, calculated as the ratio of the difference in means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered evidence of similar group means.

The table also shows the results of the equivalency-of-means tests; p-values come from a two-sample t-test, which provides evidence of whether the difference in the means is statistically significant. The equivalence test p-values are the greater of two one-sided t-test p-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. The results are used to assess the differences between the treatment and comparison groups on key characteristics likely to be associated with study outcomes.

Table B.1. Baseline characteristics of treatment and comparison groups for CCNC

Characteristics	Treatment mean (SE)	Comparison mean (SE)	Difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	68 (0.04)	69 (0.04)	-1.5 (0.06)	-2.2	-0.11	< 0.01	< 0.01
Male, %	41 (0.15)	41 (0.13)	0.79 (0.20)	1.9	0.02	< 0.01	< 0.01
Female, %	59 (0.15)	59 (0.13)	-0.79 (0.20)	-1.3	-0.02	< 0.01	< 0.01
White, %	72 (0.13)	76 (0.11)	-3.9 (0.17)	-5.3	-0.09	< 0.01	< 0.01
Black, %	24 (0.13)	20 (0.10)	4.0 (0.17)	17	0.09	< 0.01	< 0.01
Other race, %	3.1 (0.05)	3.3 (0.05)	-0.18 (0.07)	-5.6	-0.01	0.01	< 0.01
Medicare entitlement and dual eligibility status, %							
Original reason for Medicare entitlement: age	58 (0.15)	60 (0.13)	-2.0 (0.20)	-3.5	-0.04	< 0.01	< 0.01
Original reason for Medicare entitlement: disability	42 (0.15)	39 (0.13)	2.1 (0.20)	5.1	0.04	< 0.01	< 0.01
Original reason for Medicare entitlement: ESRD	0.93 (0.03)	1.0 (0.03)	-0.10 (0.04)	-11	-0.01	< 0.01	< 0.01
Dually eligible for Medicare and Medicaid	41 (0.15)	45 (0.13)	-3.8 (0.19)	-9.2	-0.08	< 0.01	< 0.01
Health status and chronic condition indicators							
HCC score[a]	1.27 (0.00)	1.40 (0.00)	-0.13 (0.00)	-10	-0.11	< 0.01	< 0.01
COPD, %	19 (0.12)	20 (0.10)	-1.2 (0.16)	-6.1	-0.03	< 0.01	< 0.01
Diabetes without complications, %	19 (0.12)	19 (0.10)	0.56 (0.16)	2.9	0.01	< 0.01	< 0.01
CHF, %	13 (0.10)	15 (0.09)	-1.7 (0.14)	-13	-0.05	< 0.01	< 0.01
Arrhythmia, %	13 (0.10)	15 (0.09)	-1.8 (0.14)	-14	-0.05	< 0.01	< 0.01

Table B.1 (*continued*)

Characteristics	Treatment mean (SE)	Comparison mean (SE)	Difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Cancers, %	11 (0.09)	11 (0.08)	0.55 (0.13)	4.9	0.02	< 0.01	< 0.01
Diabetes with acute or chronic complications, %	15 (0.11)	15 (0.09)	-0.18 (0.14)	-1.2	0.00	0.21	< 0.01
Mental health conditions, %	13 (0.10)	15 (0.09)	-1.1 (0.14)	-8.5	-0.03	< 0.01	< 0.01
Vascular disorders, %	13 (0.10)	16 (0.10)	-2.7 (0.14)	-20	-0.08	< 0.01	< 0.01
Quarter of attribution, %							
Quarter 1	59 (0.15)	60 (0.13)	-1.1 (0.20)	-1.9	-0.02	< 0.01	< 0.01
Quarter 2	8.5 (0.08)	8.0 (0.07)	0.50 (0.11)	5.9	0.02	< 0.01	< 0.01
Quarter 3	4.8 (0.06)	5.1 (0.06)	-0.30 (0.09)	-6.2	-0.01	< 0.01	< 0.01
Quarter 4	4.1 (0.06)	4.3 (0.05)	-0.16 (0.08)	-3.7	-0.01	0.06	< 0.01
Quarter 5	4.1 (0.06)	4.4 (0.05)	-0.34 (0.08)	-8.3	-0.02	< 0.01	< 0.01
Quarter 6	6.1 (0.07)	5.7 (0.06)	0.38 (0.10)	6.2	0.02	< 0.01	< 0.01
Quarter 7	3.6 (0.06)	3.4 (0.05)	0.16 (0.07)	4.4	0.01	0.03	< 0.01
Quarter 8	3.5 (0.06)	3.4 (0.05)	0.07 (0.07)	2.0	0.00	0.36	< 0.01
Quarter 9	3.8 (0.06)	3.3 (0.05)	0.47 (0.07)	12	0.02	< 0.01	< 0.01
Quarter 10	2.8 (0.05)	2.5 (0.04)	0.33 (0.06)	12	0.02	< 0.01	< 0.01
Medicare expenditures							
Total expenditures	890 (5.7)	1,023 (5.8)	-133 (8.2)	-15	-0.07	< 0.01	< 0.01

Table B.1 (*continued*)

Characteristics	Treatment mean (SE)	Comparison mean (SE)	Difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Service use							
Total hospitalizations	341 (2.9)	385 (2.6)	-44 (4.2)	-13	-0.05	< 0.01	< 0.01
Total ED or observation visits	903 (7.1)	894 (5.8)	8.6 (8.9)	< +/-1	0.00	0.34	< 0.01
Primary care visits, any setting	7,364 (23)	8,194 (23)	-811 (32)	-11	-0.11	< 0.01	< 0.01
Pharmacy factors							
Independent pharmacy, %	94 (0.07)	88 (0.08)	5.9 (0.11)	6.2	0.24	< 0.01	0.11
Rate of fills at attributed pharmacy	2.7 (0.01)	3.0 (0.01)	-0.27 (0.01)	-9.7	-0.10	< 0.01	< 0.01
Vaccines administered at pharmacy, %	90 (0.09)	59 (0.13)	30 (0.16)	34	1.00	< 0.01	1.00
Propensity score	0.52 (0.00)	0.36 (0.00)	0.15 (0.00)	29	1.03	< 0.01	1.00
Number of beneficiaries	110,968	147,034					
Omnibus test				Chi-squared statistic	Degrees of freedom	*p*-value	
				39289.93	71.00	0.00	

Source: Mathematica's analysis of CCNC-provided data on pharmacies and information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. p-values come from a two-sample t-test; equivalence test p-values are the greater of the p-values for the two one-sided t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. Unlike the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the treatment and comparison beneficiaries in this table are not weighted to account for the number of months a beneficiary was enrolled in Medicare.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CCNC = Community Care of North Carolina; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; SE = standard error.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Table C.1 displays the results from the impact analysis and Table C.2 displays the results from the sensitivity analyses. Table C.1 shows the impact estimates for the full study population, measured separately over six-month increments in intervention Years 1 and 2. The sensitivity analyses in Table C.2 shows results for the early-entry subgroup of 78,221 treatment beneficiaries attributed within the first six months of the program start date (March 1, 2015–August 31, 2015) versus the late-entry subgroup of 32,747 treatment beneficiaries attributed during the first 6 months after the launch date (September 1, 2015–May 31, 2017). Findings from the sensitivity analyses were similar to findings observed from the main impact results.

The impact estimates and sensitivity analyses models were estimated for Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the 0.10, 0.05, and 0.01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the CCNC intervention on selected Medicare FFS expenditures ($ PBPM) and use measures during one- and two-year follow-up periods

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total Medicare expenditures ($ PBPM)					
Baseline year	887	1,024			
Months 1–6	1,063	1,178	23** (10)	2.2%	0.03
Months 7–12	1,115	1,252	0.40 (11)	< 1%	0.97
Year 1	1,089	1,214	12 (9.0)	1.2%	0.19
Months 13–18	1,203	1,337	3.4 (11)	< 1%	0.75
Months 19–24	1,277	1,406	7.5 (12)	< 1%	0.52
Year 2	1,237	1,369	5.5 (9.5)	< 1%	0.56
Cumulative	1,113	1,241	10 (7.7)	< 1%	0.19
Inpatient expenditures ($ PBPM)					
Baseline year	289	322			
Months 1–6	366	390	9.5 (6.5)	2.8%	0.14
Months 7–12	398	434	-2.3 (7.0)	< 1%	0.75
Year 1	382	411	3.9 (5.6)	1.1%	0.49
Months 13–18	429	464	-1.7 (6.7)	< 1%	0.80
Months 19–24	472	498	7.3 (7.2)	2.1%	0.31
Year 2	449	479	2.6 (5.7)	< 1%	0.65
Cumulative	387	416	4.0 (4.7)	1.1%	0.40
Outpatient expenditures ($ PBPM)					
Baseline year	179	192			
Months 1-–6	202	214	1.6 (2.2)	< 1%	0.45
Months 7–12	204	218	-0.93 (2.3)	< 1%	0.69
Year 1	203	216	0.39 (1.9)	< 1%	0.84
Months 13–18	218	231	-0.15 (2.7)	< 1%	0.96

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 19–24	227	237	3.2 (2.9)	1.6%	0.27
Year 2	223	234	1.4 (2.4)	< 1%	0.56
Cumulative	211	223	0.80 (1.9)	< 1%	0.67
Hospital stays, per 1,000 beneficiaries					
Baseline year	337	383			
Months 1–6	406	445	6.3 (4.9)	1.7%	0.20
Months 7–12	430	481	-5.9 (5.1)	-1.5%	0.25
Year 1	418	463	0.47 (4.2)	< 1%	0.91
Months 13–18	460	509	-3.0 (5.3)	< 1%	0.58
Months 19–24	499	545	-0.54 (5.7)	< 1%	0.92
Year 2	478	525	-1.7 (4.6)	< 1%	0.71
Cumulative	425	470	0.51 (3.7)	< 1%	0.89
ED visits, per 1,000 beneficiaries					
Baseline year	903	895			
Months 1–6	960	941	11 (8.8)	1.2%	0.19
Months 7–12	939	955	-24** (9.4)	-2.6%	0.01
Year 1	950	948	-5.8 (7.8)	< 1%	0.46
Months 13–18	984	977	-1.0 (10)	< 1%	0.92
Months 19–24	980	982	-11 (10)	-1.2%	0.31
Year 2	983	980	-5.3 (8.9)	< 1%	0.55
Cumulative	953	950	-5.1 (7.3)	< 1%	0.48
Hospital readmissions, per 1,000 discharges					
Baseline year	0.16	0.16			
Months 1–6	0.17	0.17	0.00 (0.01)	< 1%	0.91
Months 7–12	0.19	0.18	0.00 (0.01)	1.5%	0.73
Year 1	0.18	0.17	0.00 (0.01)	1.0%	0.78
Months 13–18	0.18	0.19	-0.01 (0.01)	-4.4%	0.30
Months 19–24	0.18	0.19	-0.01 (0.01)	-4.9%	0.26
Year 2	0.18	0.19	-0.01 (0.01)	-4.7%	0.19
Cumulative	0.18	0.18	-0.00 (0.01)	-1.9%	0.53
Primary care visits, per 1,000 beneficiaries					
Baseline year	7,346	8,194			
Months 1–6	8,068	9,112	-197*** (36)	-2.4%	< 0.01
Months 7–12	8,213	9,412	-352*** (40)	-4.3%	< 0.01
Year 1	8,139	9,258	-271*** (32)	-3.3%	< 0.01
Months 13–18	8,610	9,808	-350*** (42)	-4.2%	< 0.01
Months 19–24	8,910	10,155	-397*** (45)	-4.6%	< 0.01
Year 2	8,750	9,969	-371*** (38)	-4.4%	< 0.01
Cumulative	8,315	9,463	-300*** (30)	-3.6%	< 0.01
Primary care visits, ambulatory setting, per 1,000 beneficiaries					
Baseline year	6,021	6,607			
Months 1–6	6,401	7,228	-241*** (22)	-3.7%	< 0.01
Months 7–12	6,387	7,274	-302*** (24)	-4.6%	< 0.01
Year 1	6,395	7,251	-271*** (20)	-4.1%	< 0.01
Months 13–18	6,601	7,542	-356*** (27)	-5.3%	< 0.01

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 19–24	6,726	7,720	-408*** (29)	-5.9%	< 0.01
Year 2	6,660	7,625	-380*** (25)	-5.6%	< 0.01
Cumulative	6,497	7,391	-308*** (19)	-4.6%	< 0.01
Specialty care visits, per 1,000 beneficiaries					
Baseline year	8,568	8,696			
Months 1–6	9,291	9,363	55 (44)	< 1%	0.21
Months 7–12	9,195	9,415	-92* (48)	-1.0%	0.05
Year 1	9,245	9,389	-17 (39)	< 1%	0.67
Months 13–18	9,552	9,759	-79 (51)	< 1%	0.12
Months 19–24	9,671	9,810	-12 (55)	< 1%	0.83
Year 2	9,610	9,784	-47 (46)	< 1%	0.31
Cumulative	9,250	9,411	-33 (37)	< 1%	0.36
Specialty care visits, ambulatory setting, per 1,000 beneficiaries					
Baseline year	6,677	6,621			
Months 1–6	7,281	7,271	-47* (27)	< 1%	0.08
Months 7–12	7,020	7,089	-125*** (30)	-1.8%	< 0.01
Year 1	7,154	7,184	-86*** (25)	-1.2%	< 0.01
Months 13–18	7,172	7,260	-145*** (33)	-2.0%	< 0.01
Months 19–24	6,936	6,986	-106*** (35)	-1.5%	< 0.01
Year 2	7,065	7,135	-127*** (30)	-1.8%	< 0.01
Cumulative	7,082	7,130	-104*** (24)	-1.5%	< 0.01
Percentage of beneficiaries with any ED visits in time period					
Baseline year	37	38			
Months 1–6	25	24	0.64*** (0.17)	2.7%	< 0.01
Months 7–12	24	24	0.41** (0.17)	1.7%	0.02
Year 1	37	37	0.62*** (0.19)	1.7%	< 0.01
Months 13–18	25	24	0.77*** (0.18)	3.2%	< 0.01
Months 19–24	24	23	0.67*** (0.19)	2.8%	< 0.01
Year 2	37	36	0.91*** (0.20)	2.5%	< 0.01
Cumulative	52	51	0.90*** (0.20)	1.8%	< 0.01
Percentage of beneficiaries with any hospitalization stay in time period					
Baseline year	20	23			
Months 1–6	13	12	0.60*** (0.13)	4.9%	< 0.01
Months 7–12	13	12	0.36*** (0.13)	2.9%	< 0.01
Year 1	21	20	0.67*** (0.16)	3.3%	< 0.01
Months 13–18	13	12	0.46*** (0.14)	3.9%	< 0.01
Months 19–24	13	12	0.46*** (0.15)	3.7%	< 0.01
Year 2	21	20	0.61*** (0.17)	3.0%	< 0.01
Cumulative	33	32	0.86*** (0.19)	2.7%	< 0.01
Percentage of beneficiaries with any readmission in time period					
Baseline year	3.0	3.6			
Months 1–6	2.2	1.9	0.21*** (0.06)	11%	< 0.01
Months 7–12	1.9	1.9	0.02 (0.06)	< 1%	0.77
Year 1	3.6	3.3	0.22*** (0.07)	6.6%	< 0.01
Months 13–18	1.9	1.9	-0.03 (0.06)	-1.4%	0.66

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 19–24	1.9	1.9	0.07 (0.06)	3.9%	0.23
Year 2	3.4	3.3	0.05 (0.08)	1.4%	0.54
Cumulative	6.0	5.7	0.27*** (0.10)	4.7%	< 0.01
Number of index discharges for readmissions					
Baseline year	16,712	12,435			
Months 1–6	9,749	7,142			
Months 7–12	9,138	6,943			
Year 1	18,887	14,085			
Months 13–18	8,687	6,617			
Months 19–24	7,668	5,903			
Year 2	16,355	12,520			
Cumulative	35,242	26,605			
Sample sizes					
Number of beneficiaries					
Baseline year	110,968	147,034			
Months 1–6	110,968	147,034			
Months 7–12	107,936	141,150			
Year 1	110,968	147,034			
Months 13–18	103,064	133,065			
Months 19–24	92,987	120,030			
Year 2	103,254	133,285			
Cumulative	110,968	147,034			

Source: Mathematica's analysis of information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. ED visit measures include observation stays. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate

*Significantly different from zero at the 0.10 level, two-tailed test.

**Significantly different from zero at the 0.05 level, two-tailed test.

***Significantly different from zero at the 0.01 level, two-tailed test.

CCNC = Community Care of North Carolina; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SE = standard error.

Table C.2. Sensitivity analyses: Estimated impact of the CCNC intervention on selected Medicare FFS expenditures ($ PBPM) and use measures, for beneficiaries attributed within versus after six months of program start date

	Beneficiaries attributed within 6 months of program start date					Beneficiaries attributed after 6 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total Medicare expenditures ($ PBPM)										
Baseline year	829	901				1,053	1,379			
Months 1–6	1,021	1,075	18 (11)	1.9%	0.11	1,183	1,471	38 (23)	3.6%	0.10
Months 7–12	1,080	1,164	-12 (12)	-1.2%	0.33	1,215	1,504	37 (24)	3.7%	0.12
Year 1	1,050	1,118	3.8 (9.7)	< 1%	0.69	1,198	1,488	37* (21)	3.6%	0.08
Months 13–18	1,170	1,251	-7.8 (12)	< 1%	0.53	1,294	1,580	40* (22)	3.9%	0.07
Months 19–24	1,247	1,316	3.4 (13)	< 1%	0.80	1,347	1,651	23 (25)	2.2%	0.35
Year 2	1,208	1,282	-2.0 (11)	< 1%	0.85	1,315	1,609	33* (20)	3.2%	0.10
Cumulative	1,087	1,157	2.5 (8.5)	< 1%	0.77	1,188	1,479	35** (18)	3.4%	0.04
Inpatient expenditures ($ PBPM)										
Baseline year	267	280				354	444			
Months 1–6	349	356	6.4 (7.5)	1.9%	0.39	414	484	20 (13)	5.5%	0.13
Months 7–12	387	410	-8.7 (7.6)	-2.5%	0.25	431	503	17 (16)	5.2%	0.27
Year 1	367	382	-0.85 (6.3)	< 1%	0.89	423	493	19 (12)	5.4%	0.13
Months 13–18	415	438	-9.7 (7.6)	-2.8%	0.21	468	536	23* (14)	6.7%	0.10
Months 19–24	463	473	4.2 (8.1)	1.2%	0.61	488	559	19 (15)	5.6%	0.22
Year 2	438	455	-2.7 (6.5)	< 1%	0.68	476	545	21* (12)	6.3%	0.08
Cumulative	379	393	-0.94 (5.2)	< 1%	0.86	409	479	20* (10)	5.8%	0.06
Outpatient expenditures ($ PBPM)										
Baseline year	171	181				201	223			
Months 1–6	195	200	4.8* (2.5)	2.5%	0.05	224	254	-7.6* (4.4)	-3.3%	0.08
Months 7–12	196	204	2.6 (2.6)	1.4%	0.33	224	258	-11** (4.9)	-5.0%	0.02
Year 1	196	202	3.7* (2.2)	2.0%	0.09	224	256	-9.4** (4.0)	-4.2%	0.02

Table C.2 (*continued*)

	Beneficiaries attributed within 6 months of program start date					Beneficiaries attributed after 6 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 13–18	213	219	4.1 (3.1)	2.1%	0.19	234	269	-12** (5.6)	-5.5%	0.03
Months 19–24	220	222	8.0** (3.3)	4.2%	0.01	247	282	-13** (5.9)	-5.8%	0.03
Year 2	216	221	6.0** (2.8)	3.1%	0.03	239	274	-13** (5.0)	-5.6%	0.01
Cumulative	204	209	4.8** (2.1)	2.5%	0.02	229	262	-11*** (3.9)	-4.8%	< 0.01
Hospital stays, per 1,000 beneficiaries										
Baseline year	315	343				401	498			
Months 1–6	390	413	4.3 (5.6)	1.2%	0.44	450	534	13 (9.9)	3.3%	0.18
Months 7–12	421	460	-10* (5.9)	-2.7%	0.08	455	544	8.4 (10)	2.3%	0.40
Year 1	405	436	-2.7 (4.8)	< 1%	0.57	453	539	11 (8.6)	2.8%	0.21
Months 13–18	449	482	-5.5 (6.1)	-1.5%	0.37	492	582	6.5 (11)	1.7%	0.54
Months 19–24	491	525	-5.3 (6.4)	-1.3%	0.41	513	592	18 (12)	4.9%	0.14
Year 2	469	503	-5.2 (5.3)	-1.3%	0.32	500	586	11 (9.5)	3.0%	0.25
Cumulative	418	449	-2.9 (4.2)	< 1%	0.49	445	530	12 (7.7)	3.2%	0.12
ED visits, per 1,000 beneficiaries										
Baseline year	839	821				1,086	1,109			
Months 1–6	914	884	12 (9.6)	1.3%	0.22	1,088	1,101	11 (19)	1.1%	0.56
Months 7–12	892	901	-27*** (10)	-3.1%	< 0.01	1,070	1,107	-13 (21)	-1.4%	0.52
Year 1	903	892	-7.4 (8.4)	< 1%	0.38	1,079	1,104	-0.80 (18)	< 1%	0.96
Months 13–18	963	937	7.2 (11)	< 1%	0.52	1,050	1,092	-19 (22)	-2.0%	0.40
Months 19–24	948	936	-6.2 (11)	< 1%	0.58	1,055	1,101	-23 (25)	-2.4%	0.36
Year 2	956	937	0.91 (9.6)	< 1%	0.92	1,052	1,096	-20 (21)	-2.1%	0.33
Cumulative	919	903	-2.6 (7.9)	< 1%	0.74	1,052	1,085	-9.7 (17)	< 1%	0.57
Hospital readmissions, per 1,000 discharges										
Baseline year	0.15	0.15				0.20	0.18			
Months 1–6	0.17	0.17	-0.00 (0.01)	< 1%	0.85	0.19	0.18	0.01 (0.02)	3.9%	0.64

Table C.2 (*continued*)

	Beneficiaries attributed within 6 months of program start date					Beneficiaries attributed after 6 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 7–12	0.18	0.17	0.00 (0.01)	< 1%	0.90	0.21	0.20	0.01 (0.02)	3.6%	0.66
Year 1	0.17	0.17	-0.00 (0.01)	< 1%	0.97	0.20	0.19	0.01 (0.01)	3.7%	0.59
Months 13–18	0.18	0.19	-0.01 (0.01)	-5.5%	0.26	0.20	0.20	-0.00 (0.02)	-1.3%	0.87
Months 19–24	0.18	0.19	-0.01 (0.01)	-7.2%	0.14	0.19	0.17	0.01 (0.02)	4.1%	0.68
Year 2	0.18	0.19	-0.01 (0.01)	-6.4%	0.12	0.20	0.19	0.00 (0.01)	< 1%	0.91
Cumulative	0.18	0.18	-0.01 (0.01)	-3.5%	0.33	0.20	0.19	0.00 (0.01)	2.4%	0.69
Primary care visits, per 1,000 beneficiaries										
Baseline year	7,172	7,739				7,842	9,506			
Months 1–6	7,954	8,601	-80** (39)	-1.0%	0.04	8,387	10,566	-515*** (78)	-6.0%	< 0.01
Months 7–12	8,098	8,944	-280*** (45)	-3.4%	< 0.01	8,535	10,759	-561*** (84)	-6.6%	< 0.01
Year 1	8,025	8,768	-177*** (35)	-2.2%	< 0.01	8,459	10,659	-537*** (71)	-6.3%	< 0.01
Months 13–18	8,538	9,369	-265*** (48)	-3.2%	< 0.01	8,819	11,076	-593*** (88)	-6.9%	< 0.01
Months 19–24	8,827	9,727	-333*** (51)	-3.9%	< 0.01	9,107	11,366	-595*** (100)	-6.9%	< 0.01
Year 2	8,679	9,542	-296*** (43)	-3.5%	< 0.01	8,935	11,192	-593*** (82)	-6.9%	< 0.01
Cumulative	8,243	9,029	-219*** (33)	-2.7%	< 0.01	8,518	10,716	-534*** (65)	-6.3%	< 0.01
Primary care visits, ambulatory setting, per 1,000 beneficiaries										
Baseline year	5,988	6,394				6,116	7,220			
Months 1–6	6,394	6,897	-97*** (24)	-1.5%	< 0.01	6,426	8,174	-643*** (48)	-9.3%	< 0.01
Months 7–12	6,358	6,952	-188*** (27)	-2.9%	< 0.01	6,464	8,203	-635*** (52)	-9.2%	< 0.01
Year 1	6,377	6,924	-142*** (22)	-2.2%	< 0.01	6,445	8,189	-639*** (45)	-9.2%	< 0.01
Months 13–18	6,603	7,247	-238*** (30)	-3.6%	< 0.01	6,600	8,406	-701*** (57)	-10%	< 0.01
Months 19–24	6,713	7,436	-317*** (32)	-4.6%	< 0.01	6,758	8,562	-699*** (65)	-9.9%	< 0.01
Year 2	6,657	7,339	-276*** (28)	-4.1%	< 0.01	6,664	8,470	-701*** (54)	-10.0%	< 0.01
Cumulative	6,492	7,096	-197*** (22)	-3.0%	< 0.01	6,508	8,250	-638*** (43)	-9.2%	< 0.01

Table C.2 (*continued*)

	Beneficiaries attributed within 6 months of program start date					Beneficiaries attributed after 6 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Specialty care visits, per 1,000 beneficiaries										
Baseline year	8,318	8,283				9,285	9,888			
Months 1–6	9,136	9,022	80 (49)	< 1%	0.11	9,725	10,321	7.5 (92)	< 1%	0.94
Months 7–12	9,038	9,121	-118** (54)	-1.3%	0.03	9,633	10,248	-12 (101)	< 1%	0.90
Year 1	9,089	9,071	-16 (44)	< 1%	0.71	9,682	10,289	-4.6 (84)	< 1%	0.96
Months 13–18	9,450	9,523	-107* (57)	-1.2%	0.06	9,845	10,427	22 (108)	< 1%	0.84
Months 19–24	9,539	9,532	-27 (61)	< 1%	0.66	10,000	10,547	56 (125)	< 1%	0.65
Year 2	9,496	9,528	-67 (51)	< 1%	0.19	9,911	10,480	34 (102)	< 1%	0.74
Cumulative	9,142	9,148	-41 (41)	< 1%	0.32	9,555	10,157	1.9 (80)	< 1%	0.98
Specialty care visits, ambulatory setting, per 1,000 beneficiaries										
Baseline year	6,539	6,433				7,072	7,164			
Months 1–6	7,297	7,186	4.5 (31)	< 1%	0.88	7,234	7,498	-171*** (56)	-2.3%	< 0.01
Months 7–12	7,034	7,028	-101*** (34)	-1.4%	< 0.01	6,976	7,246	-178*** (62)	-2.5%	< 0.01
Year 1	7,169	7,110	-47* (28)	< 1%	0.09	7,110	7,380	-178*** (52)	-2.5%	< 0.01
Months 13–18	7,275	7,313	-144*** (37)	-2.0%	< 0.01	6,889	7,104	-123* (69)	-1.8%	0.07
Months 19–24	6,944	6,929	-92** (38)	-1.3%	0.02	6,869	7,094	-133* (81)	-2.0%	0.10
Year 2	7,116	7,128	-118*** (33)	-1.7%	< 0.01	6,885	7,106	-129* (66)	-1.9%	0.05
Cumulative	7,117	7,090	-80*** (27)	-1.1%	< 0.01	6,980	7,231	-159*** (52)	-2.3%	< 0.01
Percentage of beneficiaries with any ED visits in time period										
Baseline year	36	36				39	42			
Months 1–6	24	24	0.74*** (0.20)	3.1%	< 0.01	26	25	0.35 (0.31)	1.4%	0.25
Months 7–12	24	23	0.30 (0.20)	1.3%	0.14	25	25	0.60* (0.31)	2.5%	0.05
Year 1	37	36	0.54** (0.22)	1.5%	0.01	38	38	0.71** (0.34)	1.9%	0.04
Months 13–18	25	24	0.99*** (0.21)	4.1%	< 0.01	24	24	0.18 (0.32)	< 1%	0.57
Months 19–24	24	24	0.81*** (0.21)	3.5%	< 0.01	23	23	0.12 (0.38)	< 1%	0.76

Table C.2 (*continued*)

	Beneficiaries attributed within 6 months of program start date					Beneficiaries attributed after 6 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Year 2	38	37	1.3*** (0.23)	3.5%	< 0.01	34	35	-0.17 (0.37)	< 1%	0.65
Cumulative	53	52	1.1*** (0.24)	2.2%	< 0.01	51	50	0.23 (0.36)	< 1%	0.52
Percentage of beneficiaries with any hospitalization stay in time period										
Baseline year	19	21				22	28			
Months 1–6	13	12	0.73*** (0.15)	6.0%	< 0.01	13	13	0.24 (0.24)	1.8%	0.32
Months 7–12	13	12	0.25 (0.16)	2.0%	0.11	13	12	0.60** (0.25)	5.0%	0.02
Year 1	21	20	0.68*** (0.19)	3.3%	< 0.01	21	21	0.58** (0.29)	2.8%	0.05
Months 13–18	13	12	0.45*** (0.16)	3.7%	< 0.01	12	12	0.46* (0.26)	3.8%	0.07
Months 19–24	13	13	0.39** (0.17)	3.1%	0.02	12	11	0.70** (0.29)	6.2%	0.02
Year 2	21	21	0.64*** (0.20)	3.1%	< 0.01	19	19	0.51 (0.32)	2.8%	0.11
Cumulative	33	32	0.95*** (0.23)	2.9%	< 0.01	31	31	0.60* (0.34)	2.0%	0.08
Percentage of beneficiaries with any readmission in time period										
Baseline year	2.8	3.2				3.6	4.7			
Months 1–6	2.1	1.8	0.23*** (0.07)	12%	< 0.01	2.3	2.2	0.16 (0.11)	7.2%	0.15
Months 7–12	1.9	1.8	0.05 (0.06)	2.5%	0.48	1.9	2.0	-0.07 (0.10)	-3.4%	0.52
Year 1	3.5	3.3	0.29*** (0.09)	8.9%	< 0.01	3.6	3.6	0.03 (0.14)	< 1%	0.83
Months 13–18	1.9	1.9	-0.05 (0.07)	-2.8%	0.43	1.9	1.9	0.04 (0.11)	2.2%	0.70
Months 19–24	2.0	1.9	0.07 (0.07)	3.9%	0.28	1.8	1.8	0.06 (0.12)	3.6%	0.61
Year 2	3.4	3.4	0.03 (0.09)	< 1%	0.77	3.1	3.0	0.11 (0.14)	3.6%	0.45
Cumulative	6.1	5.8	0.33*** (0.11)	5.8%	< 0.01	5.6	5.6	0.07 (0.17)	1.4%	0.66
Number of index discharges for readmissions										
Baseline year	12,413	9,019				4,299	3,416			
Months 1–6	7,287	5,133				2,462	2,009			
Months 7–12	6,913	5,111				2,225	1,832			
Year 1	14,200	10,244				4,687	3,841			

Table C.2 (*continued*)

	Beneficiaries attributed within 6 months of program start date					Beneficiaries attributed after 6 months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 13–18	6,662	4,971				2,025	1,646			
Months 19–24	6,102	4,669				1,566	1,234			
Year 2	12,764	9,640				3,591	2,880			
Cumulative	26,964	19,884				8,278	6,721			
Sample sizes										
Number of beneficiaries										
Baseline year	78,221	104,433				32,747	42,601			
Months 1–6	78,221	104,433				32,747	42,601			
Months 7–12	76,369	101,062				31,567	40,088			
Year 1	78,221	104,433				32,747	42,601			
Months 13–18	72,960	95,554				30,104	37,511			
Months 19–24	70,702	91,891				22,285	28,139			
Year 2	73,100	95,717				30,154	37,568			
Cumulative	78,221	104,433				32,747	42,601			

Source: Mathematica's analysis of information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary' characteristics and the probability of having any hospital stay or ED visit at baseline. ED visit measures include observation stays. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

CCNC = Community Care of North Carolina; ED = emergency department; FFS = fee for service; PBPM = per beneficiary per month; SE = standard error.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for Community Care of North Carolina (CCNC) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to CCNC. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on three core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for CCNC led to a Bayesian estimate of the program's impact on total Medicare expenditures of 1 percent (an estimated increase of $12 per beneficiary per month) in the first two program years.

Table D.1. Comparison of frequentist and Bayesian impact estimates for CCNC in the first two years after enrollment

	Impact estimate (95 percent interval)			Percentage impacts	
Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	10 (-5.1, 25)	12 (-25, 50)	3%	< 1%	1%
Hospital admissions	0.51 (-6.8, 7.8)	5.5 (-9.2, 21)	3%	< 1%	1%
ED visits	-5.1 (-19, 9.2)	7.8 (-26, 42)	2%	> -1%	< 1%

Source: Mathematica's analysis of information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

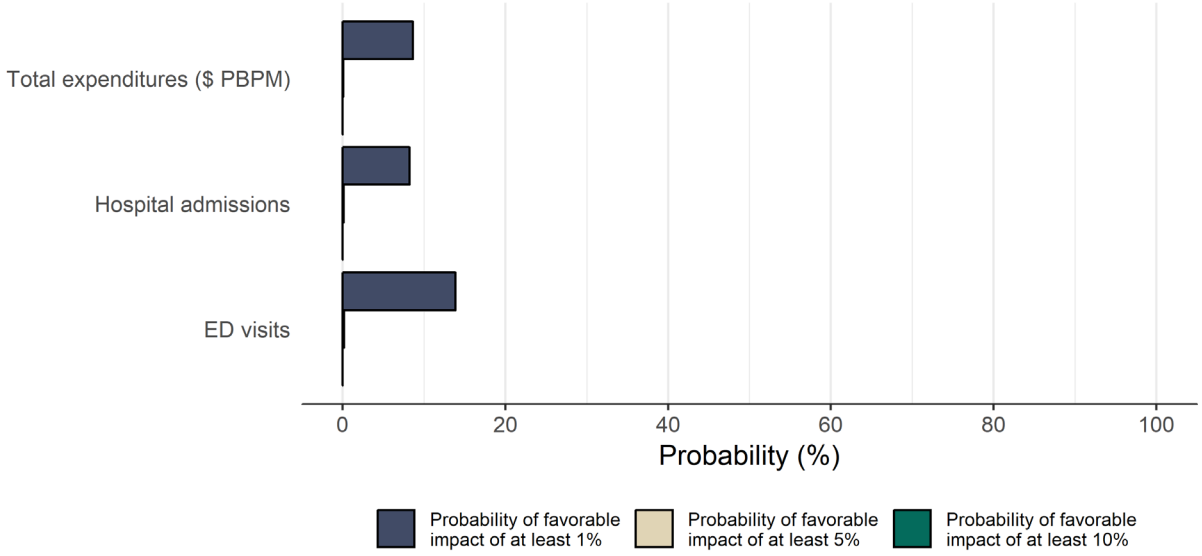
ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

In the database of HCIA R1 awardees used as prior evidence, interventions implemented in rural locations tended to have unfavorable impacts, so our prior expectations for CCNC were slightly unfavorable. The Bayesian model compromised between these expectations and the information from the frequentist CCNC evaluation to produce more neutral estimates that are more consistent across outcomes. Unusually, the Bayesian impact estimates have wider uncertainty intervals than their frequentist counterparts, likely because CCNC's enhanced pharmacy services intervention has no close correlate among the HCIA R1 awardees included in the prior. Despite these

differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that CCNC achieved favorable impacts during each of the first two years on three core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the CCNC program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from Medicare claims and enrollment data from March 2014 through August 2018, as of November 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a small probability—in the range of 10 percent—that CCNC had a favorable impact of 1 percent or more on total Medicare expenditures, hospital admissions, and emergency department visits. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the CCNC program did not have a meaningful impact on total expenditures or service utilization.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Detroit Medical Center

September 2020

James Reschovsky, Margaret O'Brien-Stain, Smita Patil, and Randall Brown

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

DETROIT MEDICAL CENTER

Detroit Medical Center (DMC) designed the Gateway to Health program to provide ongoing primary care services using a patient-centered medical home (PCMH) model to people living in Detroit, Michigan. The primary target population included individuals with diabetes, asthma, hypertension, congestive heart failure (CHF), depression, chronic obstructive pulmonary disease (COPD), or HIV infection, or frequent DMC emergency department (ED) users who did not have a primary care physician (PCP). Recruitment by patient navigators originally took place when patients visited a DMC ED. However, initially slow recruitment led DMC to expand eligibility to DMC employees and patients referred by current Gateway patients and DMC's partner organizations. The program launched in January 2015 and ended in December 2017. Table 1 summarizes the program's key characteristics.

Gateway's innovation was to locate primary care clinics within or adjacent to EDs in three Detroit hospitals operated by DMC. These clinics also had extended hours of operation. DMC leaders believed that increasing Gateway participants' access to primary care would lead to better management of chronic conditions and fewer ED visits.

Important issues for understanding the evaluation

- The program attempted to enroll patients visiting DMC's EDs who had chronic illnesses or frequent use of ED services and who lacked a PCP. The program's goal was to convince individuals to seek care from the program's primary care clinics rather than the EDs for their health problems that were not true emergencies.
- A rigorous impact analysis was not possible because of changing recruitment practices, some eligibility criteria not being available in claims data, and a low participation rate among eligible patients, indicating a serious risk of selection bias.
- Many enrolled patients never visited the Gateway clinic; many others failed to keep scheduled appointments.

After participants enrolled, they received ongoing primary care services from a multidisciplinary care team, including a behavioral health specialist, social worker, nutritionist, pharmacy educator, and a medical assistant. A nurse practitioner (NP) or a PCP led these teams. These multidisciplinary care teams enabled DMC to identify and address the full range of participants' physical, behavioral, and social needs.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	DMC designed the Gateway program to encourage frequent ED users who had no regular primary care provider and one of seven chronic conditions to use the program's primary care clinic rather than the ED for their non-emergency health care needs.
Major innovation	To make the Gateway clinics convenient to users, they were located in or adjacent to DMC hospital Eds and had extended hours of operation.
Program components	• Diversion of non-emergent ED users to Gateway clinics • Primary care medical home services • Education and training
Target population	The program sought to engage people living in Detroit and identified as frequent users of the ED who had no PCP and at least one of the following chronic conditions: diabetes, asthma, hypertension, CHF, depression, COPD, or HIV/AIDS.
Total enrollment	The program enrolled 6,996 participants, 61 percent of its original enrollment target. More than half (57 percent) of the participants were Medicaid enrollees.
Participant attrition	Nearly half of the participants did not have a single ambulatory primary care visit during the 12-month follow-up period 7 to 18 months following enrollment, suggesting low patient engagement. DMC closed the clinic at Children's Hospital of Michigan during the third year because program leaders felt the care model did not suit the needs of a pediatric patient population.
Theory of change or theory of action	DMC focused on changing participants' reliance on the ED for medical care by offering participants who sought treatment there the option of receiving immediate access to primary care at a Gateway center. The awardee hypothesized that this improved access to primary care would result in better health outcomes, fewer ED visits, and lower costs.
Award amount	$9,987,542
Effective launch date	• January 29, 2015 • The last participant enrolled on September 30, 2017.
Program setting	Gateway centers located at or adjacent to three of Detroit's largest EDs
Market area	Urban
Market location	Detroit, Michigan
Target outcomes	• Increase use of primary care office visits • Decrease ED use and total expenditures • Improve overall health among target patients
Payment model	DMC did not submit a proposal for a new payment model. Instead, it planned to continue to support the program through existing FFS billing.
Sustainability plans	DMC's parent organization, Tenet Health, signaled it would help sustain the program for another year at two of the original three locations.

CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; DMC = Detroit Medical Center; ED = emergency department; FFS = fee-for-service; PCMH = patient-centered medical home; PCP = primary care physician.

A rigorous impact evaluation of the Gateway program was not possible, for reasons provided in Table 2. Therefore, this report describes only the demographic and health characteristics of Medicaid participants at enrollment and does not present estimates of program impacts.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	A rigorous impact evaluation of this program was not possible, primarily because a key eligibility criterion (lack of a PCP) could not be replicated in Medicaid claims data. This made it impossible to develop a credible comparison group for estimating program impacts. An analysis using as the treatment group all Medicaid patients who visited a DMC ED and had one of the target conditions would have yielded unbiased estimates, but the low participation rate among eligible patients would have made detecting even large true impacts very unlikely.
Intervention group for descriptive analysis	The descriptive analysis of baseline demographic and health characteristics was conducted on 1,953 Medicaid participants who were not in managed care plans, could be identified in Medicaid claims, and had the data needed for the study, representing 49 percent of the 3,974 total Medicaid participants.
Limitations	Due to the problems noted above, the analysis cannot be used to make inferences about the impact of this program on Medicare costs or other program outcomes.

DMC = Detroit Medical Center; ED = emergency department; PCP = primary care physician.

PROGRAM DESIGN AND ADAPTATION

Overall, the program succeeded in delivering comprehensive primary care services to the population it sought to engage. Program staff and leaders felt that many participants experienced improved chronic disease management and subsequent health outcomes during their enrollment in the program.[1]

Staffing

Although the intent was for PCPs to lead Gateway clinic teams, the clinics' extended hours made recruiting physicians difficult. Program leaders adapted by making greater use of NPs as health team leaders and provided staff with appropriate training. Gateway PCMH care teams included a registered nurse care manager, certified medical assistants, behaviorists, social workers, nutritionists, and pharmacist educators. Overall, staff satisfaction was high, and turnover was infrequent, allowing there to be consistent staffing of clinics during the three-year cooperative agreement.

Facilities

Because clinics were in or adjacent to existing Eds, securing sufficient space to support operations presented a challenge over the course of the award, particularly as enrollments rose and space needs increased. Originally, DMC envisioned two clinics at Detroit Receiving Hospital (DRH), one collocated at the Rosa Parks Geriatrics Center and focused on serving an elderly population. However, space constraints at Rosa Parks forced this center to consolidate with the other DRH clinic at the end of the first program year. The opening of the clinic at Children's Hospital of Michigan was delayed during the first program year as space was

[1] The Third Annual Evaluation report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

prepared and the clinic was closed during the last program year after program leaders concluded that the model was not suited for a pediatric patient population.

Health information technology

The program made extensive use of health information technologies. DMC used its electronic medical records system to flag patients likely to meet Gateway eligibility criteria who visited the ED. DMC maintained a patient tracking system to identify successfully and unsuccessfully recruited patients. Another facilitator associated with service delivery effectiveness was the ability to use DMC's existing systems to collect self-monitoring data, access financial management data, and use a population health management tool. The utility of these data systems came under stress as a result of DMC's acquisition by Tenet Health in late 2013. Tenet's transition of DMC's information systems to its own presented challenges to program leaders when the transition disrupted functionalities, such as customizable self-monitoring reports, and slowed the timeliness of reports.

Recruiting participants

Recruiting took place when patients presented at DMC Eds. DMC flagged patients likely to be eligible in its electronic health record system based on their prior ED use or chronic condition diagnoses. Patient navigators attempted to recruit patients in the ED waiting room before their ED treatment if the triage nurse assessed that the Gateway clinic could safely treat the patient. Alternatively, navigators recruited patients as they left their ED treatment or, if admitted, social workers recruited patients during their stay for follow-up visits at Gateway. If contact at the hospital was not successful, the patient navigators continued to try to contact eligible patients by phone to recruit them.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

By far, recruitment presented the greatest challenge to program leaders. Gateway patient navigators experienced difficulties contacting ED patients and, when successful, faced high refusal and ineligibility rates resulting from patients' reports of existing relationships with PCPs. Program leaders had difficulty gaining and maintaining buy-in from ED leaders and staff throughout the life of the Gateway program. Triage nurses played a large role as they were expected to administer a medical screening examination to patients screened as low acuity on a nationally standardized Emergency Severity Index to identify patients they could safely divert to a Gateway clinic. Yet ED triage nurses at times failed to fully cooperate with Gateway patient navigators and high turnover among triage nurses required repeated training efforts.

Recruiting through the ED failed to yield the anticipated number of participants. As a result, DMC adopted supplemental recruiting strategies starting in the second program year. DMC asked other organizations, including nursing facilities and foster care agencies, to refer eligible patients to the program and opened enrollment to DMC employees. Finally, participants could

refer their family and friends to Gateway. By the end of the program in August 2017, DMC had enrolled 6,996 participants, reaching 61 percent of its original enrollment target.

In addition to problems with enrollment, the program had difficulty engaging the enrolled patients. Nearly half of the participants did not have a single ambulatory primary care visit during the 12-month follow-up period 7 to 18 months following enrollment. Gateway clinics also reported high participant no-show rates for appointments.

Implications of program implementation for achieving program goals

- Slow recruitment of participants reduced the amount of time available to observe patients' outcomes after the intervention.
- The level of patient engagement in the program was not as high as hoped, as evidenced by high no-show rates for scheduled visits and low visit rates by many participants.
- Program leaders had trouble gaining and maintaining buy-in from ED leaders and staff throughout the life of the program. High turnover of ED triage nurses created additional challenges for the program's patient navigators.
- Program staff expressed confidence that the program had positive effects on the health of enrolled patients. Program leaders reported improvements in chronic disease management and health outcomes among participants.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

The descriptive statistics relied on 1,953 adult Medicaid participants who enrolled in Gateway clinics from January 20, 2015, through September 30, 2017 (Table 3; see Table A.1, Appendix A for details on participants excluded from this analysis). Gateway patients had poor health. The risk adjustment index (Chronic Illness and Disability Payment System) that predicts Medicaid spending based on demographic characteristics and past diagnoses indicated treatment group members should have 87 percent higher expenditures than the average for Medicaid enrollees nationally. They also had high rates of respiratory, cardiovascular, psychiatric, and substance abuse conditions.

Consistent with Gateway's patients' poor health, they averaged more than three outpatient ED visits the year before enrollment and about one in five had at least one hospitalization during this period. The hospitalization rate of 471 per 1,000 patients means that those who were hospitalized tended to have multiple hospitalizations. On the other hand, Gateway's patients averaged fewer than two primary care visits during the year

Main findings from the descriptive analyses

- Gateway clinics successfully recruited the target population, that is individuals who, overall, were in poor health, frequently used hospital inpatient and ED services, and infrequently visited a primary care provider.
- The evaluation cannot make inferences about the program's impacts, due to inability to identify patients who lacked a primary care physician, and a low participation rate among all those meeting eligibility criteria observable in claims.

before enrollment—an indication of not having a PCP—and they averaged about five specialist visits (including physicians seen during ED visits) at baseline.

Table 3. Baseline characteristics of Medicaid FFS participants

Characteristic	Participants (N = 1,953)
Demographic characteristics	
Age group, %	
Younger than 25	18
25 to 34	26
35 to 44	17
45 to 54	20
55 and older	18
Male, %	50
Race, %	
White	4
Black	87
Other	8
Eligibility characteristics, %	
Managed care plan enrollee	79
Eligible due to disability	27
Medical conditions, %	
Diabetes	12
Respiratory conditions	27
Cardiovascular diseases	37
Psychiatric conditions	26
HIV and other infectious diseases	9
Substance use disorder	19
CDPS score[a]	
Mean	1.9
25th percentile	0.6
Median	1.2
75th percentile	2.4
Health care use and expenditures during the year before enrollment	
Any hospitalizations, %	20
Any outpatient ED visits, %	71
Number of hospital admissions (per 1,000 beneficiaries)	471
Number of outpatient ED visits (per 1,000 beneficiaries)	3,178
Number of ambulatory primary care visits (per 1,000 beneficiaries)	1,763
Number of ambulatory specialist visits (per 1,000 beneficiaries)	4,939
Total FFS Medicaid expenditures ($ PBPM)	$265

Sources: Mathematica's analysis of information from the awardee's finder file and Medicaid claims as of November 30, 2017.

Notes: The evaluation defined the baseline year as 4 to 15 months before each participant's enrollment date. The enrollment date is the date of a participant's first Gateway clinic visit.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid.

Table 3 (*continued*)

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

Challenges of measuring program impacts

Several factors contributed to determining it was not possible to identify a credible comparison group and conduct a rigorous evaluation of the Gateway program. First, DMC's early challenges enrolling patients in the EDs prompted program leaders to modify recruitment strategies during the three-year program period in ways that could not be replicated with claims data. Second, there was no reliable way to identify a comparison group of Medicaid enrollees who lacked a PCP. Third, those who agreed to participate likely differed in unobserved ways from those identified as eligible but who did not participate, which would lead to biased estimates of program effects. Finally, the participation rate among Medicaid ED patients who met the eligibility criteria and who could be assessed with claims data was too low for an analysis of all patients meeting the eligibility criteria observable in claims data to detect even very large true impacts on participants.

CONCLUSION

Because it was not possible to identify a strong comparison group, this study was unable to estimate the impact of the program on service use and expenditures. Many enrolled patients likely benefited from participating in Gateway, but many others did not engage sufficiently to benefit. As evidence of this, only 47 percent of the Medicaid FFS participants had a primary care visit during the 12-month follow-up period. Even in the months following the ED visit that prompted enrollment, a substantial portion failed to have a Gateway visit. Moreover, program leaders reported high no-show rates among patients with scheduled visits.

PROGRAM SUSTAINABILITY

Before its award ended in December 2017, DMC's parent organization, Tenet Health, signaled it would help sustain the Gateway to Health program for another year at two of the original three locations. However, longer-term sustainability became less certain as the awardee faced greater pressure to demonstrate that the program contained costs and improved outcomes, and program leaders and staff who did not receive assurance of ongoing employment left the health system.

DMC did not submit a proposal for a new payment model. Instead, it planned to continue to support the program through billing existing fee-for-service and transitional care management codes. To generate additional funding and to scale up the program, the awardee also worked with a consultant to reduce program costs and try to document whether the program saved money with an eye toward getting billing approval from commercial payers. It also sought to explore with payers and provider organizations how to incorporate the program into value-based payment arrangements.

This page has been left blank for double-sided copying.

Appendix A

Identifying the sample used for descriptive analysis

This page has been left blank for double-sided copying.

The analysis included only 28 percent of total participants (Table A.1). The Detroit Medical Center identified 6,996 individuals as enrolled in Gateway. The analysis was limited to adult Medicaid participants, as there were insufficient observations on participants who were enrolled in Medicare fee-for-service (FFS), Medicare Advantage, or the State Children's Health Insurance Program to support analysis. Among participants, 3,974 (57 percent) were linked to Medicaid files. Of these, the analysis dropped about half because they were in Medicaid managed care or not in Medicaid FFS for at least 90 days in the baseline year, or they did not remain enrolled in the study for at least 6 months. The final analytic sample included 1,953 Medicaid FFS participants.

Table A.1. Identifying the final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through August 31, 2017		**6,996**
Participants not enrolled in Medicaid or could not be identified in Medicaid enrollment files	3,022	3,974
Participants who did not meet enrollment and data criteria for inclusion		
Lacked 90 days of FFS enrollment during baseline period	1,698	2,276
Not enrolled 6 months past enrollment date	269	2,007
Enrolled past program end date	43	1,964
Died within 30 days of enrollment	3	1,961
Had no Medicaid coverage until quarter before enrollment in program[a]	8	1,953
Final analytic sample		**1,953**

Sources: Mathematica's analysis of information from the awardee's program encounter data base from January 2015 through September 2017 and Medicaid claims and enrollment data from July 2013 through August 2018.

[a] Because recruitment for the intervention occurred when patients visited a DMC ED, ED service use was uncharacteristically high in the quarter preceding the recruitment attempt. To avoid presenting misleading findings, the descriptive analysis excluded participants whose baseline Medicaid coverage began within the three months preceding enrollment.

DMC = Detroit Medical Center; ED = emergency department; FFS = fee-for-service.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Four Seasons Compassion for Life

September 2020

Laura Kimmey, R. Vincent Pohl, Rumin Sarwar, Bob Schmitz, and Randall Brown

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

FOUR SEASONS COMPASSION FOR LIFE

Four Seasons Compassion for Life, a nonprofit hospice and palliative care organization based in western North Carolina, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to expand the Increasing Patient and System Value with Community-Based Palliative Care (CPC) program to other providers and nearby communities. The target population consisted of Medicare fee-for-service (FFS) beneficiaries with a life-limiting illness. The goals of the program were to (1) reduce hospitalizations by 10 percent, (2) reduce in-hospital deaths by 15 percent, and (3) save more than $25 million during the three-year cooperative agreement. The HCIA R2-funded CPC program launched in September 2014. The intervention period funded by HCIA R2 ended in November 2017, after Four Seasons received a three-month no-cost extension. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The CPC program represented an expansion of an existing program and aimed to reduce hospitalizations and total expenditures among Medicare beneficiaries with life-limiting illness by providing palliative care.
- This analysis relied on 6,241 Medicare FFS beneficiaries who met the claims-based eligibility criteria. Of these, 2,097 were treatment-eligible beneficiaries who lived in one of Four Seasons' catchment areas (Henderson County) and 4,144 were comparison cases who lived in six comparison regions well-matched to Henderson County. Among the treatment-eligible group, 791 (38 percent) actually participated in CPC. These 791 participants comprised 14 percent of the 5,652 total program participants at all sites.
- A rigorous impact analysis of all participants was not possible because enrollment into the program relied heavily on clinical evaluation and judgment that could not be replicated for identifying a credible comparison group. The comparison of all eligible Medicare beneficiaries in the treatment and comparison areas eliminates this selection bias. However, other differences between the two areas rather than the program might have caused the differences in outcomes between hospitalized beneficiaries in their last year of life (38 percent of whom participated in the program) and similar beneficiaries in comparison areas. Also, the results cannot be generalized to the 86 percent of enrollees not included in the analysis.

The awardee hypothesized that palliative care received at least one year before the death of a patient with a life-limiting illness can improve the patient's quality of life and reduce the cost of health care. The CPC program provided patient-centered palliative care to participants with life-limiting illnesses through a collaborative, multidisciplinary care team that served participants' needs holistically. Services focused on achieving participants' goals related to symptom management, quality of life, psychosocial and spiritual support, coordination with community-based resources, and advance care planning. HCIA R2 funding also supported the program's activities to educate participants, families, and providers about palliative care.

This impact analysis compares changes in outcomes between a group of Medicare FFS beneficiaries in their last year of life who met program eligibility criteria assessable in claims (but were not necessarily enrolled in CPC) to outcomes for a matched comparison group. Enrollment into the CPC program depended on the provider's assessment of the beneficiary's health and prognosis. Therefore, it was not possible to identify a comparison group that would match the participants on these selection criteria and allow for a rigorous impact evaluation. To eliminate this selection bias, this analysis included only beneficiaries who died within one year of admission to a hospital or observation stay, and resided either in Henderson County, North Carolina, the location of Four Seasons' main site, or in one of the comparison regions. The six comparison regions were hospital referral regions (HRRs) that had similar demographic characteristics and end-of-life care as that used in Henderson County before CPC. The estimated differences in outcomes in the last year of life for beneficiaries in the treatment and comparison regions therefore could be due to other differences between the two areas that are unrelated to the intervention.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Four Seasons Compassion for Life enrolled patients with life-limiting illness in the CPC program and provided them with a continuum of services that addressed participants' needs and integrated their care. Four Seasons also sought to change the behavior of participants and physicians by educating participants and their families, providers, and communities about palliative care.
Major innovation	The program aimed to implement a model of community-based palliative care in inpatient and outpatient settings in health care organizations and regions other than those where Four Seasons provided palliative care before the award.
Program components	• Integrated care to deliver symptom management, social work services, disease management education, advance care planning, support with complex medical decisions, and psychosocial support • Education and training of patients and their families, physicians, and other providers
Target population	Individuals ages 65 and older who were enrolled in FFS Medicare and who had a life-limiting illness with a prognosis of surviving three years or less
Participating providers	Four Seasons; Palliative Care and Hospice of Catawba Valley; one site in Asheville, North Carolina; and two sites in Greenville, South Carolina
Total enrollment	Four Seasons enrolled a total 5,652 participants in the CPC program (73 percent of the enrollment goal).
Level of engagement	Because nearly two-thirds (63 percent) of the 5,803 participants were enrolled in the CPC program during or after the eighth program quarter; these participants might have received less exposure to program services.
Theory of change or theory of action	If a continuum of services addresses participants' needs and integrates their care in all the settings through which participants with advanced illnesses transition, the participants should have fewer hospitalizations and emergency department visits, be less likely to have an in-hospital death, and have lower total Medicare costs. If participants, families, providers, and communities are educated in palliative care, then the behavior of participants and physicians will change such that the use of community-based palliative care will increase.

Table 1 (*continued*)

Program characteristics	Description
Award amount	$9,569,123
Effective launch date	September 2, 2014
Program settings	Any setting in which a participant received health care, including specialty care clinics, hospitals, long-term care facilities, hospices, primary care practices, and a participant's private residence
Market area	Rural, suburban, urban
Market location	Western North Carolina and Greenville, South Carolina
Target outcomes	• 10 percent reduction in hospitalizations for CPC participants • 15 percent fewer in-hospital deaths among CPC participants • $25,272,000 in total Medicare savings on the cost of care for participants who receive the CPC intervention during the three-year cooperative agreement
Payment model	New Medicare FFS payment, bundled, or episode payment
Sustainability plans	Continuing the program unchanged from the award period with funding from health insurers billed for services; developing a new capitated payment model

CPC = Community-Based Palliative Care program; FFS = fee-for-service.

The impact analysis presented in this report included 2,097 Medicare FFS beneficiaries who lived in one of Four Seasons' catchment areas (Henderson County), had a hospital stay, and died within one year of that hospital admission. This treatment-eligible group included 791 CPC participants (38 percent). The study identified a comparison group of 4,144 Medicare FFS beneficiaries who met the same criteria but resided in six HRRs that were similar to Henderson County but were not in CPC's catchment area. Table 2 summarizes the key features of the evaluation. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on two different models, depending on the outcome: (1) a first-differences model using the difference between the outcome of interest during the follow-up period and the baseline year as the dependent variable for continuous outcomes; and (2) a post-period comparison of outcomes between eligible treatment beneficiaries and the comparison group for binary outcomes. Both models used regression analysis to control for differences in baseline characteristics that might be correlated with outcomes.
Intervention group for evaluation	The treatment group for this analysis included 2,097 Medicare FFS beneficiaries who lived in Henderson County, North Carolina, and were likely eligible to participate in the CPC program (that is, they had a hospitalization and died within one year after that admission). Among these 2,097 beneficiaries, the CPC program enrolled 791 (38 percent). The intervention group included only Henderson County residents because the participation rate among eligible beneficiaries in other counties was too low to support an evaluation.

Table 2 (*continued*)

Features	Description
Comparison group	The comparison group included 4,144 beneficiaries who met the same criteria and lived in one of six HRRs with pre-program demographic characteristics and end-of-life care similar to that of Henderson County before CPC.
Limitations	If treatment-eligible beneficiaries differed from the comparison group in ways not captured in Medicare administrative files and claims, the impact estimates might be biased. More importantly, other factors in the treatment and comparison areas unrelated to the intervention might have affected outcomes differently for patients in the two areas. The 38 percent participation rate among the treatment-eligible group means that impacts on those actually receiving the intervention are likely to be about 2.5 times larger than the estimates obtained on the treatment-eligible group. This analysis might not detect even large true effects (for example, 20 percent) on participants.

CPC = Community-Based Palliative Care program; FFS = fee-for-service; HRR = hospital referral region.

PROGRAM DESIGN AND ADAPTATION

Four Season's CPC program had two components: (1) integrated palliative care and (2) patient education and provider training.[1] The analysis could not measure the independent effect of each intervention component on changes in outcomes.

Integrated care

The CPC program did not undergo major changes to the type of health care services provided during the cooperative agreement. The awardee sought to address participants' needs holistically—for example, by providing spiritual and social support as well as clinical care. The highly collaborative, multidisciplinary CPC care teams integrated inpatient and outpatient care such that it spanned all settings through which participants with advanced illnesses transition, such as hospitals, clinics, private residences, nursing homes, and assisted living facilities. A nurse practitioner or a physician assistant oversaw the care teams; they oversaw registered nurses, social workers, and administrative support staff.

According to the program's protocols, CPC care teams were to schedule in-person home appointments within 48 hours of enrollment for high-risk participants; for low-risk participants, the care teams were to schedule a home visit within 7 to 10 days or an in-person clinic visit within two weeks of enrollment. Care teams followed up with participants in person or by phone as needed throughout the remainder of their enrollment. During the first encounter, program staff typically assessed the participant's health, developed a care plan with input from the participant and caregivers, and documented decisions for advance care planning. Other services included symptom management, social work, education in disease management, support with complex medical decisions, and psychosocial support.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Education and training

The awardee trained CPC providers and referring providers about how to judge whether to refer patients to the program based on the primary diagnosis, physical limitations, prognosis, and other elements listed in a paper screening tool developed by Four Seasons. The training consisted of a 40-hour immersion course on palliative care, cultural competency, and other relevant topics, along with ongoing training to implementation sites through weekly or monthly calls, summary of quality monitoring, and one-on-one communication with members of the team as needed. The awardee also offered participant and family education about palliative care.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

A review of qualitative and quantitative information suggests that Four Seasons successfully implemented the CPC program. According to data it submitted, the awardee hired and retained staff throughout the cooperative agreement despite challenges in the general palliative care field with workforce shortages and staff burnout. Four Seasons received positive feedback about the training it provided to program staff and clinicians. One respondent to a staff survey described the ongoing training from Four Seasons as "solid and continuous" support that helped the implementing sites to "focus on what we need to improve."

Implications of program implementation for detecting impacts

- Program participants received comprehensive services that addressed their needs holistically, including spiritual and social support and symptom management.
- There were fewer participants than the awardee had anticipated who were most likely to benefit from the CPC program because they were either seriously ill or were transitioning from one type of care to another. This led to a suboptimal patient mix and might have limited the program's ability to reduce expenditures.
- The awardee was confident, however, that the program had achieved its intended goals.

However, Four Seasons faced several challenges implementing its program. First, the awardee had to revise its program enrollment target after finding that many more patients than expected were ineligible to participate in the CPC program because they were enrolled in Medicare Advantage plans. Second, misperceptions about palliative care among participants and their families also created challenges in enrollment and service delivery. However, the efforts of Four Seasons and its implementing partners to expand its community outreach efforts by launching a patient and family education module succeeded in overcoming these misperceptions, according to interview respondents. Third, due to the nature of the CPC program, many participants were enrolled for only a short time before death. This happened most likely because providers who referred these beneficiaries to Four Seasons' program did so only shortly before the participants' death. One-quarter of treatment group

members who participated in the program died within 20 days of enrollment and 50 percent died within 70 days. Because of short enrollment periods, the intervention might have had a smaller effect than the awardee had expected.

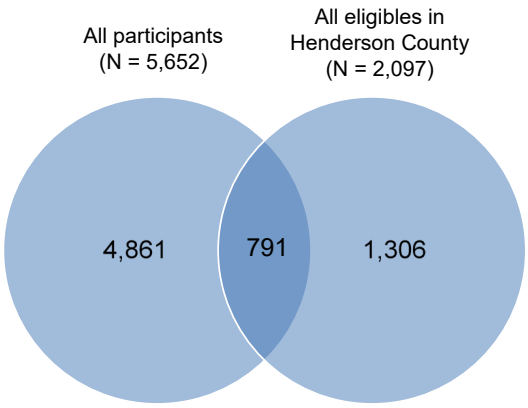
ESTIMATING PROGRAM IMPACTS

Study sample

Because Medicare claims data do not contain the type of clinical information providers used to identify beneficiaries to enroll in the program, it was not feasible to identify a comparison group that matched well to the CPC program participants. Enrollment into the CPC program relied heavily on the provider's assessment of the beneficiary's health and prognosis. Elements that providers considered when assessing health and prognosis included whether the provider would be surprised if the patient died in the next three years; physical limitations such as fall risk; presence of serious illness, such as an advanced or end-stage disease; and social determinants such as housing status, substance abuse, and lack of caregiver support. Most of these data are not available in Medicare claims.

To minimize the risk of bias due to self-selection into palliative care, the treatment group used for this analysis consisted only of beneficiaries who met certain criteria (described below) that made it likely they were eligible to enroll in the program. The analysis further restricted the treatment group to beneficiaries who lived in Henderson County, where Four Seasons has its main location, because a high proportion of those meeting these evaluation criteria actually did enroll in CPC. To be included in the analysis, treatment group members had to meet four criteria: (1) in Medicare FFS in the enrollment month and at least three months in the year before enrollment, (2) had at least one hospital admission in the year before enrollment, (3) died within one year of the last hospital admission in the year before enrollment, and (4) were not in hospice in the 90 days before enrollment. The potential comparison group included all Medicare beneficiaries who met the same four criteria and lived in one of six HRRs in which health care use by Medicare beneficiaries in the last two years of life was similar to that of beneficiaries in Henderson County before the program start. This approach ensured that the treatment and comparison groups had comparable access to palliative care before the program began.

Because members of the comparison group and non-enrolled members of the treatment group did not have an enrollment date, and participants often enrolled some time after the hospital discharge, the evaluation assigned pseudo-enrollment dates to these sample members. The pseudo-enrollment date for a given non-enrolled eligible treatment group member or comparison group member at an assigned number of days after the index hospital discharge date. The assigned number of days was randomly selected so that the distribution of days between the index hospital discharge date and the pseudo-enrollment date for these sample members matched the distribution of actual time between hospital discharge and enrollment dates for program participants. (For two-thirds of the participants, the enrollment date occurred during a hospitalization.)



Participants enrolled in the CPC program on a rolling basis from September 2014 to August 2017. Among the total of 5,652 FFS Medicare beneficiaries who participated in the program, the analysis included only the 791 (14 percent) who lived in Henderson County. Because Four Seasons is based in Henderson County and had the highest participation rate there, the evaluation restricted the analysis to this area. The treatment group included an additional 1,306 Medicare beneficiaries not enrolled in the program but who met the selection criteria described earlier. Hence, the program enrolled 38 percent of the eligible treatment group. The evaluation used propensity score matching to select the comparison group, and it consisted of 4,144 Medicare FFS beneficiaries. (Appendix A, Table A.1 describes the identification of the analytic sample).

Characteristics of treatment and comparison group beneficiaries

The treatment and comparison groups were similar in terms of demographic characteristics, expected future health care costs, and health care use and expenditures in the baseline year (Table 3). Most beneficiaries in both groups were 75 or older, and they were predominantly White. Following the sample selection criteria, all beneficiaries had a hospital admission (or observation stay; the table shows the percentage with an admission) during the year before enrollment and more than half of the sample had an emergency department (ED) visit during the year before enrollment. Their disease burden was relatively high, with an average hierarchical condition category (HCC) score of almost 4 in the treatment and comparison groups, meaning the study sample had expected annual Medicare costs four times the national average of all FFS beneficiaries. Almost 25 percent of beneficiaries had an HCC score of 5 or higher. Average spending per beneficiary per month (PBPM) was $3,044 for the treatment group and $3,034 for the comparison group, more than twice the national average. Although the treatment and comparison group areas had similar measures of end-of-life care for chronically ill Medicare beneficiaries, there might have been unmeasured differences, and those differences might have been related to study outcomes.

Due to the sample selection criteria, all beneficiaries died within one year of their last hospital admission before enrollment. As a result, the follow-up period lasted at most 12 months, and it was typically much shorter. Treatment beneficiaries died within 78 days on average after enrollment and comparison beneficiaries died within 113 days of their pseudo-enrollment date on average. The 35-day difference in survival after enrollment between the treatment and comparison beneficiaries might suggest differences in end-of-life care between Henderson County and the comparison regions. Alternatively, the timing of enrollment for program participants might relate to specific changes in their health or services received, whereas no such association exists for members of the treatment or comparison groups who did not enroll in the

program. Appendix B provides the full balance results measured during the 12 months before the enrollment date.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Measure	Treatment (N = 2,097)	Comparison (N = 4,144)
Demographics, %		
Age group		
65 to 74	21	16
75 to 84	33	39
85 and older	46	45
Male	46	46
Race, %		
White	97	98
Black	2.4	1.2
Other	1.0	1.0
Original reason for Medicare eligibility, %		
Old age and survivor's insurance	88	88
Disability insurance benefits	11	11
End-stage renal disease	0.4	0.2
Medicare and Medicaid dual status, %		
Not dually eligible	83	83
Dually eligible	17	17
HCC score[a]		
Mean	3.9	3.8
25th percentile	2.5	2.4
Median	3.6	3.6
75th percentile	5.0	4.9
Baseline expenditures ($ PBPM)		
Total expenditures	3,030	3,023
Service use during the year before enrollment, %		
Hospital stay[b]	96	95
ED visit	54	61

Sources: Mathematica's analysis of information from awardee's finder file as of November 30, 2017, and Medicare claims and enrollment data as of August 31, 2019.

Notes: The baseline period covers the 12-month period before the enrollment (or pseudo-enrollment) date that led to sample inclusion for each beneficiary.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. Observations on comparison beneficiaries were also weighted to reflect the number of different treatment group beneficiaries to which the comparison beneficiary was matched.

Appendix B presents full balance results. Exact matching variables include the index date of hospital discharge.

Table 3 (*continued*)

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. The analysis used the most recently available HCC algorithms to calculate HCC scores.

[b] All sample members had to have had either an inpatient or an observation stay during the year before enrollment or pseudo-enrollment. Almost all had an inpatient stay.

ED = emergency department; FFS = fee-for-service; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact analysis relied on two different models, depending on the outcome: (1) a first-differences model for continuous outcome measures, using the difference between the outcome of interest during the follow-up period and the baseline year as the dependent variable; and (2) a post-period comparison of binary outcomes between eligible treatment beneficiaries and a set of matched comparison beneficiaries. Both models controlled for differences in baseline characteristics that might be correlated with outcomes. The follow-up period was the time from the enrollment (or pseudo-enrollment) date to the date of death. The study outcomes included Medicare expenditures PBPM and measures of health care use for up to one year after enrollment, as well as expenditures and hospital use during the last 7, 14, and 30 days of participants' lives. For all beneficiaries, the study calculated measures of total expenditures and expenditure categories as the difference between the outcome during the follow-up period and the baseline year. Appendix A contains a detailed description of the study sample, the statistical models, and the outcomes used to estimate the treatment–comparison differences.

IMPACT RESULTS

The study found that expenditures were an estimated 10 percent higher among treatment group beneficiaries than among comparison group beneficiaries, not lower as expected (Table 4). The higher hospice and skilled nursing facility (SNF) expenditures for the treatment group than the comparison group drove the higher total spending for the treatment group (see Appendix C). Although the rate of hospice use did not differ significantly, the analysis estimated hospice spending among beneficiaries in the treatment group to be on average 59 percent higher than among comparison group members. Appendix C presents the full results of the impact analysis. Appendix D shows the results from the Bayesian analysis.

Table 4. Estimated impacts of the Four Seasons Compassion for Life intervention on selected outcomes during a 12-month follow-up period

	Treatment-comparison difference	Percentage change in outcomes[a]	*p*-value
Total Medicare expenditures ($ PBPM)	601*	10%	0.08
Hospice expenditures ($ PBPM)	480***	59%	< 0.01
Percentage with a hospital admission	-8.1	-10%	0.26
Percentage with in-hospital death	-1.1	-12%	0.41

Table 4 (*continued*)

Sources: Mathematica's analysis of information from awardee's finder file as of November 30, 2017, and Medicare claims and enrollment data as of August 31, 2019.

Notes: Due to the approach used to select the sample, all beneficiaries died within the first 12 months of the follow-up period. Appendix C presents full impact estimates. Appendix D shows the results from the Bayesian analysis.

[a] Percentage difference is equal to the ratio of the estimated difference divided by the treatment group mean minus the estimated difference.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per beneficiary per month.

These findings suggest treatment group members might have received more intensive or longer hospice services, though the study cannot make that claim with confidence because there was a substantial risk of unobserved differences between the comparison and treatment groups. The higher SNF expenditures and lower home health expenditures for the treatment group (shown in Appendix C) might suggest differences between treatment and comparison geographic areas in sources of post-acute care, or to the timing of enrollment for program participants. In addition, Four Seasons' experience delivering palliative care and the fact that care teams focused on symptom management and psychosocial needs might have led to an increased recognition of care needs better addressed in inpatient hospice or SNF settings.

Main findings from impact evaluation

- Due to limitations in the research design, the findings from this analysis might not be reliable measures of program impacts.
- The treatment group had total Medicare expenditures that exceeded those of the comparison group by $601 PBPM.
- Higher hospice and SNF expenditures drove the higher total expenditure.
- Estimated effects on both hospitalization and in-hospital death rates were both favorable, but were not statistically significant.

The estimated difference between treatment and comparison groups might be associated with program participation. For example, hospice expenditures were higher among treatment group members who participated in Four Seasons' program than among treatment group members who did not participate in the program, and program participants were more likely to use hospice than nonparticipants in the treatment group. The awardee's four decades of experience providing hospice care in Henderson County might explain the more intensive use of hospice services. It is possible that Four Seasons could better identify the needs of its enrollees and provided more extensive and more expensive hospice services than hospice agencies serving non-enrolled treatment beneficiaries in the same area. The evaluation did not estimate differences between enrollees and non-enrolled treatment beneficiaries in hospice use.

The treatment group had lower hospitalization rates and in-hospital death rates, but the association of treatment status with these outcomes was not statistically significant. The likely failure to substantially reduce the percentage of beneficiaries hospitalized probably explains why

total expenditures did not decline as the awardee had hoped. The higher hospice and SNF expenditures for the treatment group were not offset by an equally large or larger reduction in inpatient expenditures. The awardee might not have engaged participants early enough to reduce hospitalizations and overall expenditures. Most enrollees participated in the program for a brief time before their death. Despite its efforts to educate patients and families about palliative care, Four Seasons might not have had enough time to help participants substitute lower-cost palliative care for hospitalizations.

CONCLUSIONS

Although the awardee successfully enrolled participants in its CPC program and delivered palliative care services to them as intended, this analysis suggests that the intervention did not reduce Medicare spending among patients near the end of life in Henderson County. Total expenditures were higher for the treatment group than for the comparison group due to higher hospice and SNF spending. The findings point to a more intensive use of hospice care among the participants of Four Seasons' program. Due to its extensive experience with delivering hospice services and the focus during implementation on educating of patients, families, and providers, it is possible that the awardee identified unmet care needs among program participants. Delivering these services might have led to an associated increase in total expenditures. At the same time, there was no compelling evidence that the CPC program led to fewer hospitalizations, which contradicts the theory of action. The findings suggest, however, that the awardee did not achieve its goal of saving $25 million during the three-year cooperative agreement. Although the estimates suggest that CPC did not save money, the estimates do suggest there might have been a modest improvement in patients' experience among patients near the end of life in Henderson County. Because patients often express a preference for dying at home instead of in an inpatient setting, the estimated reduction in in-hospital death rates likely had positive implications for their quality of life. This is in line with the program's theory of action.

Limitations of evaluation

The palliative care intervention could have caused the estimated differences in outcomes between the eligible treatment group and the matched comparison group. However, the estimated differences in outcomes could be due to other factors affecting the patterns of end-of-life health care use and expenditures in Henderson County and the HRR comparison areas. In addition, the program enrolled only 38 percent of the beneficiaries in the treatment group, which substantially dilutes estimated program effects on eligible participants. This analysis includes only 14 percent of all participants in Four Seasons' CPC program during the period covered by the cooperative agreement. It is possible that the program had different impacts on health care use and expenditures for most enrollees, whom the analysis could not include because of the low participation rate among seemingly eligible patients in those geographic areas.

PROGRAM SUSTAINABILITY

After its award ended in November 2017, Four Seasons reported that all five participating sites continued the CPC program without major changes. Four Seasons had always anticipated that it would continue the CPC program beyond the award period because it had operated the program at a single site for 12 years before the award. The awardee partnered with four additional sites to implement the program during the award period, all of which continued the program with one change: Four Seasons no longer oversaw the partners' programs, which also meant that the sites no longer reported data to Four Seasons or received program-related data feedback reports from Four Seasons.

Four Season's proposed payment model

Four Seasons proposed paying for the CPC program through a bundled payment model. The PBPM payments would cover the following standard set of palliative care and hospice services:

- Advance care planning
- Up to three goals-of-care conferences
- Home visits
- Clinic visits
- Symptom management
- Coordination of services
- Social work
- Some services provided by the hospice team

Services unrelated to palliative care were carved out of the payment model, including hospitalizations, primary care, and specialty care. The awardee partnered with the American Academy of Hospice and Palliative Medicine to develop the model.

Four Seasons and its implementing partners sustained CPC at their sites by funding the program the same way Four Seasons did before the award—billing insurers when possible and using internal funding or external grants to cover the rest of the program costs. Knowing that these funding streams could not reliably sustain CPC in the long term, Four Seasons continued to work on securing funding for its bundled payment model after the award ended. The awardee submitted an alternative payment model to the Physician-Focused Payment Model Technical Advisory Committee, which the committee approved for a demonstration according to the awardee. The awardee was also in talks with commercial payers to fund its proposed capitated payment model but had not reached any agreements as of July 2018.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

Sample construction

Treatment group

The treatment group for the analysis consisted of Medicare fee-for-service (FFS) beneficiaries who satisfied the following conditions: (1) lived in Henderson County, North Carolina, where Four Seasons operates its main site, during the admission month; (2) had a hospital or observation stay with an admission date from September 1, 2013, to August 31, 2017; (3) died within one year from the admission date; (4) were enrolled in Medicare FFS during the month of the discharge from the hospitalization or end of the observation stay; and (5) were at least 65 years old. These selection criteria identified 2,385 beneficiaries in claims data who met these analytic criteria for inclusion in the treatment group, of whom 820 were already enrolled in Four Seasons' Community-Based Palliative Care (CPC) program. Some beneficiaries had multiple hospitalizations that met those criteria. For beneficiaries enrolled in the CPC program, the evaluation used the most recent admission before the enrollment date as the index admission for each beneficiary. For non-enrolled beneficiaries with multiple admissions during the program period, the evaluation treated a randomly selected admission as the index admission.

Table A.1 shows the how the evaluation defined the participant portion of the analysis sample for this study. The table lists why the evaluation excluded participants and the number of participants withdrawn for each reason.

Table A.1. Number of participants excluded from impact analysis, by reason

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through August 31, 2017		**5,652**
Excluded beneficiaries who:		
Were not found in Medicare crosswalk file	159	5,493
Did not live in Henderson County, North Carolina	3,762	1,731
Did not have any inpatient or observation stay with an admission date from September 1, 2013 to August 31, 2017	133	1,598
Did not die within one year of admission	571	1,027
Were enrolled in FFS in the month of discharge	3	1,024
Were younger than 65 during the month of admission	1	1,023
Were not living in treatment geographic area during month of admission	37	986
Did not have any inpatient or observation stay claim during year before enrollment date on the finder file	166	820
Were in hospice at any time during the 90 days before the index hospitalization	29	791
Final analytic sample		**791**

Source: Mathematica's analysis of information from awardee's finder file as of November 30, 2017, and Medicare claims and enrollment data as of August 31, 2019.

FFS = fee-for-service.

Pseudo-enrollment date

The definition of the pseudo-enrollment date differs between enrolled and not enrolled beneficiaries in Four Seasons' CPC program. For beneficiaries enrolled in the program, the evaluation defined the pseudo-enrollment date as the program enrollment date if enrollment did not occur on the admission date of the qualifying hospitalization or observation stay. If enrollment occurred on the date of the hospital admission or start of the observation stay, the evaluation defined the index date as one day after the enrollment date. This ensures that the costs associated with the qualifying hospital or observation stay are part of the baseline period and not the intervention period, because the program typically identified beneficiaries as candidates for CPC after they arrived at the hospital.

For members of the treatment group not enrolled in the program, the evaluation assigned the pseudo-enrollment date by adding to the index hospitalization date a number drawn randomly from the distribution of days from hospital admission to the actual enrollment date for the enrolled beneficiaries. This ensured that the distribution of the number of days from hospital admission to the pseudo-enrollment date for both the comparison group and the non-enrolled treatment group members matched the distribution of days from hospital admission to actual enrollment date of the enrolled beneficiaries (for example, the program participants).

The pseudo-enrollment date defined the baseline and follow-up periods, which are both beneficiary specific. The baseline comprised the 365 days before the pseudo-enrollment date. The follow-up period started on the pseudo-enrollment date and varied in length because it lasted from the pseudo-enrollment date to each beneficiary's death. Because of the sample selection criteria, the follow-up period lasted at most 365 days.

To arrive at the final treatment group used in the study, the analysis dropped the following observations on non-enrolled treatment group members (matching the criteria used to select program participants included in the analysis): the pseudo-enrollment date occurred on the admission date of a subsequent hospitalization; the beneficiary was in hospice within 90 days before the pseudo-enrollment date; the beneficiary was not enrolled in Medicare FFS in the month of the pseudo-enrollment date or at least three months in the year before the pseudo-enrollment date; the beneficiary did not live in Henderson County in the month of the pseudo-enrollment date; or the beneficiary died on or before the assigned pseudo-enrollment date. The final treatment group consisted of 2,097 beneficiaries, of whom 791 were enrolled in the program.

Potential comparison group

The potential comparison group consisted of Medicare FFS beneficiaries who satisfied the same conditions as treatment group members, except that they lived in one of six comparison hospital referral regions (HRRs) instead of Henderson County during the month of the qualifying hospital admission or observation stay. These selection criteria identified 60,003 potential comparison group members.

Comparison geographic areas

Because local conditions influence health care use and expenditures, drawing the comparison group from areas adjacent to the awardee's service area would be ideal. However, this was not feasible, because organizations that referred patients to the awardee treated many qualifying beneficiaries. The awardee worked with many hospitals, long-term care facilities, skilled nursing facilities, clinics, and other organizations to obtain referrals for CPC. Because many of these organizations were located in counties near Henderson County, the evaluation could not select a potential comparison group that excluded all beneficiaries treated by these organizations using administrative data.

The analysis used a two-step process to select comparison geographic areas whose end-of-life care and demographic characteristics were similar to Henderson County. The analysis sample consisted of beneficiaries in the last year of life. Therefore, in Step 1, the analysis used the 2013 Dartmouth Atlas to find HRRs where end-of-life care was similar to the Asheville, North Carolina, HRR, where Henderson County is located, based on three measures of health care use by chronically ill Medicare beneficiaries at the end of life:

1. Percentage of decedents receiving hospice benefits
2. Percentage of deaths occurring in hospital
3. Hospital care intensity index, which is a standardized ratio of inpatient days to inpatient admissions

The study examined data on these three measures to identify HRRs that had values similar to the Asheville HRR. For example, 54.8 percent of chronically ill Medicare beneficiaries who died in 2013 in the Asheville HRR received hospice benefits. Therefore, the study limited potential comparison HRRs to those where the rate of hospice use ranged from 49.8 to 59.8 percent. Other HRRs in North Carolina and the southeastern states have markedly different patterns of end-of-life care than the Asheville HRR; those HRRs have lower rates of hospice use, higher rates of in-hospital death, and higher values of the hospital care intensity index than Asheville. Therefore, the potential comparison geographic areas identified in Step 1 are in other regions of the country.

In Step 2, the study examined potential comparison HRRs identified in Step 1 to determine which of them had counties with somewhat similar demographic characteristics to Henderson County based on the following county-level measures from the Area Resource File:

- Percentage urban population, 2010
- Median household income, 2013
- Percentage Black or African American or Hispanic, 2013

Based on data on health care use at the end of life and demographic characteristics, the study selected the following HRRs as comparison areas: Iowa City, Iowa; Waterloo, Iowa; Portland,

Maine; Muskegon, Michigan; Petoskey, Michigan; and Salem, Oregon. The evaluation refers to them as the comparison geographic areas.

Pseudo-enrollment date

For potential comparison group members, the evaluation defined the possible pseudo-enrollment date as 1, 2, 3, 5, 7, 10, 14, 30, or 90 days after the admission date of the qualifying hospital stay if the resulting index date was at least one day before the beneficiary's death. That is, each beneficiary who lived in one of the comparison geographic areas and met sample eligibility criteria had up to nine versions in the potential comparison group, and each version had a different pseudo-enrollment date. The analysis selected this distribution of days (1, 2, 3, and so on) based on examining the distribution of days from admission to enrollment among beneficiaries in the treatment group who enrolled in the CPC program. From these, up to nine possible observations per beneficiary, the evaluation excluded observations for the following reasons: the pseudo-enrollment date occurred on the admission date of a subsequent hospitalization; the beneficiary was in hospice or had a claim for palliative care within 90 days before the pseudo-enrollment date; the beneficiary was not enrolled in Medicare FFS in the month of the pseudo-enrollment date or at least three months in the year before the pseudo-enrollment date; the beneficiary did not live in a comparison geographic area in the month of the pseudo-enrollment date; or the beneficiary died on the pseudo-enrollment date. The potential comparison group used for propensity score matching had 333,601 observations—each with a unique combination of beneficiary identification number and pseudo-enrollment date.

Description of modeling strategy and outcome variables

The analysis estimated program impacts on total and service-specific expenditures using a first-difference approach. Specifically, it subtracted expenditures per beneficiary per month (PBPM) in the baseline period (the 12-month period before a beneficiary's index date) from expenditures PBPM in the follow-up period (the period from the index date to the beneficiary's death, which was always less than one year after the index date). The estimates show the regression-adjusted change between baseline and intervention periods for the treatment group relative to that for the comparison group. These regressions control for beneficiaries' characteristics and number of hospital stays, emergency department (ED) visits or observation stays, and primary care visits during the baseline period. The evaluation then regression-adjusted treatment–comparison differences of estimates for the binary outcomes of any hospital stay and any ED visit during the follow-up period based on regressions that controlled for a beneficiary's baseline characteristics and whether the beneficiary had any hospital stay and any ED visit, respectively, during the baseline period. The regressions for any hospital stay and any ED visit also controlled for the beneficiary's number of hospital stays, ED visits or observation stays, and primary care visits and total expenditures per month during the baseline period.

In addition to the standard outcomes described in Appendix A of Volume I of this report, awardee-specific outcomes included total expenditures in the last 7, 14, and 30 days of life and binary outcomes of any hospital stay in the last 30 days of life; any hospice stay during the

follow-up period; and in-hospital death. The analysis calculated total expenditures in the last 7, 14, and 30 days of life only for beneficiaries who survived for at least those numbers of days during the follow-up period. The analysis calculated the outcome of any hospital stay in the last 30 days of life only for beneficiaries who survived for at least 30 days during the follow-up period. The outcome equals one if a beneficiary was admitted to a hospital within 30 days of the date of death. The outcome of any hospice stay equals one if the beneficiary used any hospice services in the follow-up period.

The regressions did not control for a program maturity indicator. At the start of the Round 2 of the Health Care Innovation Award grant period, Four Seasons had provided outpatient palliative care in Henderson County for 12 years. Four Seasons did not substantially change the palliative care services it offered or its selection of patients during the award period. The only major change was that the program began collecting and monitoring patients' data through its Quality Data Collection Tool database. Its collection and use of patients' data changed throughout the three-year award period, and the analysis identified no particular time when collection and use of those data matured.

To account for different lengths of time observed, the analysis weighted regressions for the outcomes of total and service specific expenditures PBPM, any hospital stay, any ED visit, and any hospice stay by the number of days from the index date to a beneficiary's death.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. The table displays the weighted means of baseline characteristics for the 2,097 treatment beneficiaries and the 4,144 matched comparison beneficiaries used in the analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable, calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables include demographic characteristics (age, gender, and race); Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category [HCC] score and chronic condition indicators); Medicare expenditures in total and by type of service; and service use. The analysis required an exact match on whether the beneficiary was hospitalized on the index date.[1] The analysis measured variables over various specified intervals before each beneficiary's index date.

The table also shows the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence of whether the difference in the means is statistically significant. The equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values equivalence tests, which assess whether the treatment and comparison group means differ by more than 0.25 standard deviations. Finally, the study also performed an omnibus test in which the null hypothesis is that the treatment and matched comparison groups balanced across all linear combinations of the covariates. The results assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely associated with study outcomes.

Final sample

The study selected the comparison group by propensity score matching. The final estimation sample consisted of 6,241 beneficiaries—2,097 treatment group members and 4,144 comparison group members. The treatment group included 791 beneficiaries who participated in Four Seasons' Community-Based Palliative Care (CPC) program during the funding period, met the selection criteria used for this evaluation, and resided in Henderson County, North Carolina, plus 1,306 Medicare fee-for-service (FFS) nonparticipating beneficiaries who resided in Henderson County and met the same sample selection criteria. The comparison group consisted of 4,144 Medicare beneficiaries who resided in six comparison hospital referral regions, met the sample selection criteria, and were selected as matches for the treatment group members. Although the potential comparison group had multiple observations per beneficiary, the matched comparison group included only one observation for a given comparison beneficiary.

The two groups matched well on most characteristics, but a few differences caused concerns about possible bias. Proportionate differences that are large are not necessarily a concern if they are due to small absolute differences between the groups divided by a small mean (for example, the difference in hospice care during the baseline period is only 1.5 percentage points, but 41

[1] The index date is the first day of the post-period. The measure of whether a beneficiary was hospitalized on his or her index date reflects whether the beneficiary was in the hospital or discharged from the hospital on the index date.

percent of the mean). However, differences such as those observed for the proportion with dementia with complications (13 versus 8.5 percent) and the proportion with depressive disorder (12 versus 8.6 percent) exceeded 30 percent of the treatment group mean.

Table B.1. Baseline characteristics of treatment and matched comparison groups for FSCL

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Demographics							
Age, years	82 (0.19)	83 (0.12)	-0.28 (0.25)	< +/-1	-0.03	0.27	< 0.01
Male, %	46 (1.1)	46 (0.77)	-0.21 (1.6)	< +/-1	0.00	0.89	< 0.01
White, %	97 (0.40)	98 (0.23)	-1.2 (0.53)	-1.3	-0.07	0.02	< 0.01
Black, %	2.4 (0.33)	1.2 (0.17)	1.2 (0.43)	51	0.08	< 0.01	< 0.01
American Indian, Alaska Native, Asian or Pacific Island American, or other, %	0.48 (0.15)	0.81 (0.14)	-0.33 (0.25)	-70	-0.05	0.19	< 0.01
Hispanic, %	0.19 (0.10)	0.05 (0.03)	0.14 (0.10)	75	0.03	0.16	< 0.01
Unknown, %	0.33 (0.13)	0.14 (0.06)	0.19 (0.15)	57	0.03	0.21	< 0.01
Dual eligibility status, %							
Dually eligible for Medicare and Medicaid	17 (0.82)	17 (0.57)	0.19 (1.2)	1.1	0.01	0.87	< 0.01
Original reason for Medicare eligibility, %							
Old age and survivor's insurance	88 (0.70)	88 (0.49)	-0.24 (1.0)	< +/-1	-0.01	0.81	< 0.01
Disability insurance benefits	11 (0.69)	11 (0.49)	0.02 (0.98)	< +/-1	0.00	0.98	< 0.01
End-stage renal disease	0.43 (0.14)	0.21 (0.07)	0.21 (0.17)	50	0.03	0.21	< 0.01
Health status and diagnoses							
HCC score[a]	3.9 (0.04)	3.8 (0.03)	0.10 (0.06)	2.6	0.05	0.08	< 0.01
Acute renal failure, %	41 (1.1)	41 (0.76)	0.48 (1.5)	1.2	0.01	0.76	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Anemia, %	41 (1.1)	42 (0.77)	-0.91 (1.5)	-2.2	-0.02	0.55	< 0.01
CHF, %	55 (1.1)	54 (0.77)	0.69 (1.5)	1.3	0.01	0.65	< 0.01
COPD, %	33 (1.0)	33 (0.73)	-0.14 (1.5)	< +/-1	0.00	0.92	< 0.01
Dementia with complications, %	13 (0.73)	8.5 (0.43)	4.2 (0.95)	33	0.13	< 0.01	< 0.01
Dementia without complications, %	29 (0.99)	26 (0.68)	2.5 (1.4)	8.6	0.05	0.08	< 0.01
Diabetes with acute complications, %	0.91 (0.21)	1.2 (0.17)	-0.33 (0.32)	-37	-0.04	0.30	< 0.01
Electrolytes, %	62 (1.1)	61 (0.76)	0.79 (1.5)	1.3	0.02	0.59	< 0.01
Major depressive disorder, %	12 (0.72)	8.6 (0.43)	3.6 (0.98)	30	0.11	< 0.01	< 0.01
Metastatic cancer and acute leukemia, %	13 (0.73)	14 (0.54)	-0.91 (1.0)	-7.0	-0.03	0.37	< 0.01
Morbid obesity, %	7.9 (0.59)	6.3 (0.37)	1.6 (0.78)	20	0.06	0.04	< 0.01
Protein-calorie malnutrition, %	38 (1.1)	36 (0.74)	1.9 (1.5)	5.1	0.04	0.20	< 0.01
Septicemia, %	32 (1.0)	26 (0.68)	6.2 (1.4)	19	0.13	< 0.01	< 0.01
Vascular disease, %	28 (0.99)	35 (0.74)	-6.5 (1.4)	-23	-0.14	< 0.01	< 0.01
Medicare expenditures							
Total expenditures[b]	2,951 (53)	2,981 (34)	-31 (73)	-1.0	-0.01	0.68	< 0.01
Total expenditures, 3 months before enrollment	6,978 (148)	7,451 (100)	-472 (208)	-6.8	-0.07	0.02	< 0.01
Acute inpatient expenditures[b]	1,318 (27)	1,429 (19)	-111 (39)	-8.4	-0.09	< 0.01	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Service use							
Total hospitalizations	1,635 (25)	1,597 (16)	38 (35)	2.3	0.03	0.27	< 0.01
Total hospitalizations, 3 months before enrollment	4,654 (57)	4,626 (38)	29 (79)	< +/-1	0.01	0.71	< 0.01
Total ED or observation visits	1,106 (33)	1,141 (21)	-35 (43)	-3.2	-0.02	0.42	< 0.01
Total ED or observation visits, 3 months before enrollment	1,925 (76)	2,193 (49)	-268 (101)	-14	-0.08	< 0.01	< 0.01
Primary care visits, any setting	14,896 (251)	12,684 (140)	2,212 (323)	15	0.19	< 0.01	0.02
Primary care visits, any setting, 3 months before enrollment	26,196 (469)	23,622 (283)	2,574 (628)	9.8	0.12	< 0.01	< 0.01
Hospice use in baseline, %[c]	3.8 (0.42)	2.2 (0.23)	1.5 (0.52)	41	0.08	< 0.01	< 0.01
Reasons for sample inclusion, %							
Categorical days from most recent hospitalization to index date, Category 1	34 (1.0)	38 (0.76)	-4.6 (1.5)	-14	-0.10	< 0.01	< 0.01
Categorical days from most recent hospitalization to index date, Category 2	22 (0.90)	19 (0.60)	3.3 (1.3)	15	0.08	0.01	< 0.01
Categorical days from most recent hospitalization to index date, Category 3	44 (1.1)	43 (0.77)	1.4 (1.5)	3.1	0.03	0.38	< 0.01
Qualified due to observation stay	6.8 (0.55)	10 (0.47)	-3.6 (0.87)	-52	-0.14	< 0.01	< 0.01
No clinician visit or hospital stay on index date	27 (0.97)	27 (0.69)	-0.14 (1.3)	< +/-1	0.00	0.92	< 0.01
Clinician visit on index date	8.6 (0.61)	8.5 (0.43)	0.14 (0.83)	1.7	0.01	0.86	< 0.01
Propensity score	0.38 (0.00)	0.37 (0.00)	0.01 (0.01)	2.1	0.04	0.20	< 0.01
Number of beneficiaries	2,097	4,144					

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Omnibus test				Chi-squared statistic 473.34	Degrees of freedom 40.00	*p*-value 0.00	

Sources: Mathematica's analysis of information from awardee's finder file as of November 30, 2017, and Medicare claims and enrollment data as of August 31, 2019.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The analysis calculated the comparison group means in the table by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix A, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. The numbers in this table differ slightly from those in Table 3 in the report, due to the use of follow-up period weights in constructing the means presented there. Those weights were equal to the proportion of the follow-up period observed. Exact matching variables include index date during hospitalization.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. The analysis used the most recently available HCC algorithms to calculate the HCC scores.

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

[c] The hospice measure used for matching includes use of hospice on the day of enrollment. This has little bearing on the matches selected. For estimating impacts on hospice, the follow-up period outcome measure of hospice use includes admission to hospice on the day of enrollment (or pseudo-enrollment).

CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; FSCL = Four Seasons Compassion for Life; HCC = hierarchical condition category; SE = standard error.

Appendix C

Detailed results from impact analyses

This page has been left blank for double-sided copying.

Table C.1 displays the results from the analysis for the full sample of 6,241 beneficiaries. The analysis estimated models for Medicare expenditures and probability of using any service, in total and by type of service. The estimated percentage change in outcomes is the estimated change in outcomes divided by a counterfactual value defined as the treatment group mean minus the estimated change in outcomes. One, two, and three asterisks indicate estimated changes in outcomes that differ statistically from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated changes in select Medicare FFS expenditures (dollars PBPM) and use measures associated with the FSCL intervention during a 12-month follow-up period

	All beneficiaries					
	Treatment group mean	Comparison group mean	Estimated change in outcomes (SE)	Percentage change in outcomes[a]	Participation rate[b]	*p*-value
Total Medicare expenditures ($ PBPM)						
Baseline year	3,030	3,023				
12-month follow-up period	6,441	6,337	601* (344)	10%	0.38	0.08
Total Medicare expenditures ($ PBPM)[d]						
Baseline year	2,976	3,005				
12-month follow-up period	5,850	5,658	634** (258)	12%	0.38	0.01
Acute inpatient expenditures ($ PBPM)						
Baseline year	1,391	1,468				
12-month follow-up period	1,853	2,191	-34 (210)	-1.8%	0.38	0.87
Acute inpatient expenditures ($ PBPM)[d]						
Baseline year	1,329	1,437				
12-month follow-up period	1,452	1,704	-25 (138)	-1.7%	0.38	0.86
Other inpatient expenditures ($ PBPM)						
Baseline year	121	123				
12-month follow-up period	111	157	-66 (79)	-37%	0.38	0.41
Hospital outpatient expenditures ($ PBPM)						
Baseline year	469	449				
12-month follow-up period	555	639	-148 (103)	-21%	0.38	0.15
Professional Part B expenditures ($ PBPM)						
Baseline year	442	455				
12-month follow-up period	915	843	102 (65)	13%	0.38	0.12

Table C.1 (*continued*)

	All beneficiaries					
	Treatment group mean	Comparison group mean	Estimated change in outcomes (SE)	Percentage change in outcomes[a]	Participation rate[b]	*p*-value
Home health expenditures ($ PBPM)						
Baseline year	136	133				
12-month follow-up period	220	244	-59** (26)	-21%	0.38	0.03
SNF expenditures ($ PBPM)						
Baseline year	424	339				
12-month follow-up period	1,451	1,306	353*** (120)	32%	0.38	< 0.01
Durable medical equipment expenditures ($ PBPM)						
Baseline year	39	45				
12-month follow-up period	43	64	-26* (15)	-38%	0.38	0.08
Hospice expenditures ($ PBPM)						
Baseline year	7.4	7.9				
12-month follow-up period	1,293	893	480*** (83)	59%	0.38	< 0.01
Percentage of beneficiaries with any hospital admission in a time period						
Baseline year	96	95				
12-month follow-up period	74	77	-8.1 (7.2)	-9.9%	0.38	0.26
Percentage of beneficiaries with any ED or observation visits in a time period						
Baseline year	54	61				
12-month follow-up period	65	72	-4.5 (3.3)	-6.4%	0.38	0.18
Percentage of beneficiaries with hospice use in a time period						
Baseline year	0.81	0.55				
12-month follow-up period	62	59	2.5 (2.3)	4.2%	0.38	0.28
Total Medicare expenditures during the last 7 days of life ($ PBPM)[e]						
12-month follow-up period	2,727	2,057	670*** (259)	33%	0.38	< 0.01
Total Medicare expenditures during the last 14 days of life ($ PBPM)[f]						
12-month follow-up period	5,056	4,285	771** (384)	18%	0.37	0.04

Table C.1 (*continued*)

	All beneficiaries					
	Treatment group mean	Comparison group mean	Estimated change in outcomes (SE)	Percentage change in outcomes[a]	Participation rate[b]	*p*-value
Total Medicare expenditures during the last 30 days of life ($ PBPM)[g]						
12-month follow-up period	8,938	9,069	-131 (644)	-1.4%	0.37	0.84
Total Medicare expenditures during the last 7 days of life ($ PBPM)[d,e]						
12-month follow-up period	2,483	1,975	508*** (195)	26%	0.38	< 0.01
Total Medicare expenditures during the last 14 days of life ($ PBPM)[d,f]						
12-month follow-up period	4,738	4,083	656** (320)	16%	0.37	0.04
Total Medicare expenditures during the last 30 days of life ($ PBPM)[d,g]						
12-month follow-up period	8,640	8,715	-75 (540)	< 1%	0.37	0.89
Percentage with hospital stay in the last 30 days of life[g]						
12-month follow-up period	33	37	-4.7* (2.8)	-13%	0.37	0.09
Percentage with an in-hospital death						
12-month follow-up period	8.1	9.2	-1.1 (1.4)	-12%	0.38	0.41
Sample sizes						
Number of beneficiaries						
Baseline year	2,097	4,144				
12-month follow-up period	2,097	4,144				
Survived at least 7 days of the follow-up period	1,685	3,770				
Survived at least 14 days of the follow-up period	1,394	3,437				
Survived at least 30 days of the follow-up period	1,093	2,929				

Sources: Mathematica's analysis of information from awardee's finder file as of November 30, 2017, and Medicare claims and enrollment data as of August 31, 2019.

Note: Estimates effects on expenditures PBPM during the 12-month follow-up period relied on a first-difference approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and follow-up periods. The estimate for the binary outcomes of any hospital stay, ED visit, or hospice use is a regression-adjusted treatment–comparison difference based on a regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The estimate for outcomes in the last 7, 14, or 30 days of life is a regression-adjusted treatment–comparison difference based on a regression that controls for a beneficiary's characteristics. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment or pseudo-enrollment.

Table C.1 (*continued*)

[a] Percentage change in outcomes is relative to a counterfactual value defined as the treatment mean minus the estimated change in outcomes.

[b] The participation rate is the number of participants among treatment group beneficiaries—that is, those who actually received the intervention—divided by the total number of treatment group beneficiaries who were eligible to receive the intervention.

[c] The adjusted change in outcomes represents the estimated effect of the intervention on only the participants—that is, those who received the intervention. It is derived by dividing the estimated change in outcomes for all eligible treatment group beneficiaries by the participation rate.

[d] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

[e] Sample includes only beneficiaries who survived at least 7 days of the follow-up period.

[f] Sample includes only beneficiaries who survived at least 14 days of the follow-up period.

[g] Sample includes only beneficiaries who survived at least 30 days of the follow-up period.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; FSCL = Four Seasons Compassion for Life; PBPM = per beneficiary per month, SE = standard error, SNF = skilled nursing facility.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the evaluation also estimated the program impacts for Four Seasons Compassion for Life (FSCL) using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. Drawing probabilistic conclusions requires external or prior evidence. In this analysis, the findings from the evaluation of 87 awardees included in Round 1of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to FSCL. The evaluation calculated probabilities using the results of a Bayesian regression that models FSCL's impacts on total Medicare expenditures jointly with impacts from HCIA, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimate for total Medicare expenditures with the regression estimate obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimate from the frequentist regression for FSCL led to a Bayesian estimate of the program's impact on total Medicare expenditures of 9 percent (an estimated increase of $248 per beneficiary per month) during the year after enrollment.

Table D.1. Comparing frequentist and Bayesian impact estimates for FSCL in the first year after enrollment

	Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	601 (-73, 1,275)	464 (-63, 995)	7%	10%	8%

Source: Mathematica's analysis of information from awardee's finder file as of November 30, 2017, and Medicare claims and enrollment data as of August 31, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; FSCL = Four Seasons Compassion for Life; HCIA R1 = Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results are imprecise, the Bayesian model gave more weight to the prior and produced somewhat more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that FSCL's impact on total Medicare expenditures is statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impact corresponds to a high probability of achieving policy targets, such as a 5 percent reduction in expenditures. For FSCL, there is less than a 3 percent probability of achieving a 1 percent

reduction in total Medicare expenditures, reaffirming the frequentist findings that the program did not meaningfully reduce costs.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: The Fund for Public Health in New York

September 2020

Cordon Newhart, Jim Derzon, Kyle Emery, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

THE FUND FOR PUBLIC HEALTH IN NEW YORK

The Fund for Public Health in New York (FPHNY) received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create and support the implementation of Project INSPIRE. The goal of the program was to improve uptake of and adherence to hepatitis C virus (HCV) drug treatment. The target population consisted of Medicare and Medicaid enrollees, 18 years of age or older with a detectable HCV ribonucleic acid (RNA) viral load and who had difficulty keeping appointments, had received sporadic care, had never been in care, or who requested support and lived in New York City. The Project INSPIRE program launched January 2015, three months after the award. The intervention period funded by HCIA R2 ended in February 2018. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program aimed to improve adherence to drug treatment; reduce unnecessary hospitalizations, readmissions, and ED visits; and reduce costs for beneficiaries in New York City through care coordination, tele-mentoring and peer navigation for adults with HCV.
- The impact analysis relied on 327 Medicare and 1,310 Medicaid fee-for-service (FFS) beneficiaries who enrolled in the program and 1,275 Medicare and 5,085 Medicaid comparison group beneficiaries with HCV who did not receive care at either of the two participating health systems.
- Because adverse health effects from HCV can develop decades after contracting the virus, it is likely that some of the impact of Project INSPIRE will occur after the follow-up period used in the analysis.
- Selection bias could influence the results of the study because beneficiaries who enrolled in the program might have had poorer health status and/or a higher motivation to engage in treatment than comparison group beneficiaries.

Awardee leaders hypothesized that Project INSPIRE could lower Medicare and Medicaid spending and reduce unnecessary hospitalizations, readmissions, and emergency department (ED) visits by (1) preventing HCV infection from advancing to hepatocellular carcinoma or other forms of liver disease; (2) stabilizing and managing participants during the program; (3) improving participants' self-sufficiency by facilitating the treatment and management of comorbid conditions, such as HIV, substance abuse and mental health disorders, diabetes, and heart disease; and (4) preventing repeated treatment for HCV by avoiding treatment failures and reinfections. Care coordinators and peer navigators helped participants complete HCV treatment by addressing their underlying health problems that commonly interfere with adherence—including mental health and substance abuse issues—while teaching participants the skills they need to manage their health independently. Project INSPIRE also provided tele-mentoring for primary care and other providers to increase their capacity to treat HCV.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The program sough to facilitate HCV treatment for participants by improving clinical and nonclinical care for both HCV and comorbid conditions and by using tele-mentoring to increase the capacity of health care providers to effectively treat HCV.
Major innovation	The program used care coordination to guide an underserved population through HCV treatment by addressing the underlying health problems that commonly interfere with adherence to treatment and by using tele-mentoring to enable primary care providers to treat HCV.
Program components	• Care coordination to administer psychosocial assessments, present health promotion modules, and provide social and instrumental support to promote treatment adherence • Tele-mentoring for primary care providers to increase awareness of HCV treatment and the HCV population and to provide continuing medical education credits facilitating engagement and the certification of primary care providers as approved prescribers of HCV drug therapy in New York State • Peer navigation to facilitate an apprehensive populations' engagement with Project INSPIRE by sharing similar experiences
Target population	The program sough to engage HCV-positive individuals with a detectable HCV RNA viral load who were born from 1945 to 1965; who resided in the Bronx or in East or Central Harlem in New York City; who were eligible for Medicare or Medicaid; and who had difficulty keeping appointments, had received sporadic care, had never been in care, or who had requested support. Other adults ages 18 and older who lived in the five boroughs of New York City could also participate, although they were not actively recruited. According to Medicaid and Medicare claims data, there were about 11,000 Medicare and 85,000 Medicaid beneficiaries with HCV that visited Mount Sinai or Montefiore during the study period that would have been eligible to participate in the program.
Participating providers	Mount Sinai's 8 hospitals, 9 hospital affiliates, and 29 network-affiliated physician practices and Montefiore's 10 hospitals and more than 200 outpatient ambulatory care sites participated in the program.
Total enrollment	A total of 2,775 beneficiaries (945 Medicare and 1830 Medicaid) enrolled in the program.
Level of engagement	According to awardee data, among the participants who were clinically suitable for treatment, 82 percent initiated treatment. However, not all enrollees completed the six-week drug treatment program, as 76 percent were adherent. The high caseloads of care coordinators likely reduced participants' engagement. More than one-third of respondents (37 percent) to a staff survey reported that staffing was a minor or major barrier to effective service delivery.
Theory of change/ theory of action	Care coordinators identified patients with psychosocial needs who would benefit from the program. After enrolling patients, care coordinators performed outreach to engage participants and performed a range of care management activities, including calling participants, sending letters, providing MetroCards for transportation, and developing relationships with them. The awardee expected the patient-centered service model to improve HCV cure rates by increasing access to treatment, coordinating care and health promotion services, and addressing comorbid conditions that can interfere with HCV treatment. These outputs were, in turn, expected to lead to better health, appropriate care use, and lower costs.
Award amount	$9,948,459
Effective launch date	January 15, 2015
Program settings	Health systems (the Mount Sinai Health Network and Montefiore Health System, Inc.)
Market location	New York City
Market area	Urban

Table 1 (*continued*)

Program characteristics	Description
Target outcomes	• Increased adherence to HCV drug therapy • Increased SVR or HCV cure rate • Improved participant satisfaction • Reduction in episodes of acute care for behavioral conditions
Payment model	One-time bundled payment from Medicaid MCOs to fund care coordination services for patients with HCV; program services would cost an estimate $760 per beneficiary during the study period
Sustainability plans	While awaiting full acceptance and implementation of its payment model, the awardee and its hospital partners obtained a commitment and resources from hospital leadership to continue the program in the short term and were trying to obtain additional funding at the end of the award. They also modified the program to contain its costs while making it more robust by reducing the frequency of tele-mentoring for providers and updating health promotion education materials to be more engaging and less time-consuming.

HCV = hepatitis C virus; RNA = ribonucleic acid; MCO = managed care organization; SVR = sustained viral response.

While 2775 beneficiaires enrolled in the program, the impact analysis relied on 327 Medicare FFS beneficiaries and 1,310 Medicaid beneficiaries who enrolled in Project INSPIRE from January 2015 through February 2017, six months before the end of the award, met program eligibility criteria and had complete Medicaid or Medicare data (see Table 2 below and Tables A.1 and A.2 in Appendix A for details on sample exclusions). The comparison groups consisted of 1,275 Medicare and 5,085 Medicaid beneficiaries that met program eligibility criteria (but who did not visit one of the participating health systems during the enrollment period, and thus were unlikely to have been recruited for the program. Table 2 summarizes the key features of the evaluation.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group.
Intervention group for evaluation	There were 2,775 beneficiaries (945 Medicare and 1830 Medicaid) that enrolled in the program. However, 585 Medicare beneficiaries were excluded from the sample for the impact analysis because they did not meet Medicare eligibility requirements (mainly because they were not in Medicare fee-for-service) and 33 Medicare beneficiaries were excluded because they did not meet the program eligibility requirements (ages 18 and older with HCV living in New York City and visiting Montefiore or Mount Sinai). Likewise, 106 Medicaid beneficiaries were excluded because they had incomplete Medicaid data and 412 Medicaid beneficiaries were excluded because they did not meet program eligbility requirements.After these exclusions, there were 327 Medicare and 1,310 Medicaid FFS beneficiaries included in the impact analsysis who enrolled in Project INSPIRE from January 2015 through February 2017.
Comparison group	The comparison groups consisted of 1,275 Medicare and 5,085 Medicaid FFS beneficiaries who met the same criteria as the treatment group, except they did not visit Montefiore or Mount Sinai during the baseline period (and therefore were not likely to have been recruited by the program).

Table 2 (*continued*)

Features	Description
Limitations	First, two health care systems, which could differ from other hospitals or systems in the services they offer patients with HCV, implemented the intervention. Thus, results might not be generalizable to other settings. Second, the release of a new class of drugs to treat HCV with fewer adverse side effects coincided with better enrollment than expected and those who enrolled might have had greater motivation to initiate and adhere to HCV drug treatment. Finally, many of the benefits of treating HCV occur many years after treatment, after the end of the evaluation period.

FFS = fee for service; HCV = hepatitis C virus.

PROGRAM DESIGN AND ADAPTATION

The Project INSPIRE service delivery model had three components: (1) care coordination, (2) tele-mentoring, and (3) peer navigation.[1]

Care coordination

Care coordinators at both clinical partners administered a psychosocial assessment to identify participants who met established thresholds for psychosocial intervention. They enrolled and administered the appropriate health promotion modules, and arranged for mental health services, substance abuse services (required, if needed, to authorize drug treatment for HCV), and other services to address comorbid conditions associated with HCV. Participants received a multitude of care coordination services, including setting up and reminding participants about medical appointments, completing insurance authorizations, offering coaching and services promoting participant health, establishing linkages to clinical and non-clinical providers, and providing other forms of social and instrumental support that helped participants adhere to treatment. After participants completed treatment, care coordinators made reminder calls to encourage them to return for SVR testing 12 weeks after treatment.

Tele-mentoring

Project INSPIRE provided tele-mentoring for primary care and other providers to increase their capacity to treat HCV and engaged providers by facilitating weekly training and the exchange of information between hepatologists and program providers. Tele-mentoring increased the familiarity of primary care providers with HCV, treatment, and the HCV-afflicted population. Tele-mentoring also offered continuing medical education credits, which facilitated provider engagement and the certification of primary care providers as approved prescribers of HCV drug therapy in New York State. In Year 2, tele-mentoring sessions evolved from teaching sessions to case conferencing, including presentations and group discussions of HCV cases.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Peer navigation

Peer navigators facilitated participant engagement with an apprehensive population by sharing similar experiences, encouraging the participants to stay engaged, and meeting participants' needs. Navigators often accompanied participants to appointments, provided MetroCards, and placed reminder calls about medication adherence.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The possibility of curing a disease facilitated providers' involvement and improved care. Project leadership reported that providers "really love" the idea of curing a deadly disease, something providers might not often have the chance to do. Payment also increased providers' involvement by enabling providers to join tele-mentoring sessions and engage with the intervention. Training how to effectively treat HCV was not readily available elsewhere and Project INSPIRE's funding enabled providers to attend the HCV tele-mentoring training sessions to learn how to effectively treat patients with HCV. Providers were successful in getting participants to initiate and adhere to treatment. Among the 2,218 participants who were clinically suitable for treatment, 82 percent initiated treatment, 76 percent were adherent to treatment, and 70 percent obtained a sustained viral response (SVR).

Implications of program implementation for detecting impacts

- Tele-mentoring facilitated providers' engagement in the program and improved fidelity to effective service delivery.
- Although care coordinators were generally effective in outreach and enrolling participants, their heavy caseloads might have impeded their ability to effectively engage enrollees.
- Selection bias was a concern for this study because beneficiaries who enrolled might have had higher motivation to engage in treatment or poorer health status at baseline than those who chose not to enroll. Those who did not enroll might have been waiting for the new class of highly effective HCV drugs that had fewer adverse side effects than previously available drugs. On the other hand, those who enrolled might have had poorer health status and worse outcomes regardless of any program effects. Because of multiple possible sources of selection bias, this study could have over- or understated true program effects.

However, engaging participants became more challenging as care coordinators' caseloads grew. In interviews, respondents noted that when coordinators reached the originally planned 125-participant caseload, they did not have enough time to do all the work necessary to engage participants and deliver the full range of intervention services. About one-third of respondents (32 percent) to the staff survey reported that staffing was a minor barrier, and 5 percent reported that staffing was a major barrier to effective service delivery. Another such barrier was that some participants might not have seen the value in care coordination. Participants working multiple jobs said they wanted to "get in and get out" of their

provider's office. This impeded the ability of care coordinators to effectively engage participants in ongoing care management.

ESTIMATING PROGRAM IMPACTS

Study sample

This study relied on 327 Medicare and 1,310 Medicaid FFS beneficiaries who enrolled in Project INSPIRE from January 2015 through February 2017. Treatment group beneficiaries had to meet the following eligibility criteria: be at least 18 years old, reside in New York City, have an HCV diagnosis during the baseline period, and visit Montefiore or Mount Sinai. The comparison groups consisted of 1,275 Medicare and 5,085 Medicaid FFS beneficiaries who met the same criteria as the treatment group, except they did not visit Montefiore or Mount Sinai during the baseline period (and therefore were not likely to have been recruited by the program). (Appendix A, Tables A.1 and A.2 describe the identification of the analytic sample for Medicare and Medicaid, respectively).

Characteristics of treatment and comparison group beneficiaries

Comparing treatment and comparison group characteristics at baseline confirmed that the two groups were well balanced (see Appendix B for full matching results) and the Medicare and Medicaid beneficiaries shared enough characteristics to be pooled for analysis (Table 3). Most participants were ages 45 to 64; male; and of Black, Hispanic, or unknown race. About 75 percent of the Medicare sample members were dually eligible for Medicaid.

Consistent with the program's eligibility criteria, many treatment group members had chronic conditions related to HCV and, because they were sick, they were frequent users of health care services. Nearly 20 percent of treatment group Medicaid beneficiaries and 33 percent of treatment group Medicare beneficiaries had cirrhosis of the liver. About 18 percent of the Medicare treatment group and 12 percent of the Medicaid treatment group had filled an HCV prescription in the two years before enrollment. About 33 percent of Medicaid and 40 percent of Medicare participants had a hospitalization in the 12-month baseline period. In that same period, nearly 50 percent of the treatment group had experienced an outpatient ED visit. In sum, the enrolled population required more services and was likely more expensive to treat than the general Medicaid or Medicare population.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Measure	Medicaid beneficiaries		Medicare beneficiaries	
	Treatment (N = 1,310)	Comparison (N = 5,085)	Treatment (N = 327)	Comparison (N = 1,275)
Demographics				
Age at enrollment, years	52	53	64	65
Age group, %				
18 to 44 years	17.4	17.5	2.9	4.0
45 to 54 years	34.0	33.0	11.0	10.0
55 to 64 years	47.0	47.0	28.0	26.0
65 to 74 years	1.4	1.7	48.0	44.0
75 or older	0.2	0.9	10.0	16.0
Race group, %[a]				
White, %	1.4	3.0	32.0	33.0
Black, %	2.5	3.8	49.0	49.0
Hispanic/Latino, %	48.0	45.0	14.0	13.0
Asian, %	0.2	0.7	2.8	4.2
Other race, %	0.0	0.1	1.8	0.0
Unknown race, %	48.0	47.0	1.8	0.6
Male, %	65.0	65.0	68.0	65.0
Dually eligible for Medicaid, %[b]	n.a.	n.a.	75	74
Health status during year before enrollment				
CDPS score[c]	4.5	4.5	n.a.	n.a.
HCC score[d]			2.2	2.1
Percentage with cirrhosis	19.0	17.0	33.0	29.0
Percentage with liver transplant	0.7	0.6	2.1	2.3
Percentage with hepatocellular carcinoma	1.6	1.5	4.6	3.5
Service use and expenditures during the year before enrollment				
Any hospitalizations, %	33.0	33.0	40.0	37.0
Any outpatient ED visits, %	47.0	46.0	42.0	44.0
Any hepatitis C prescription fill 2 years before enrollment	12.0	11.0	18.0	16.0
Total Medicare expenditures ($ PBPM)	n.a.	n.a.	2,931	2,756

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicare and Medicaid FFS claims, encounter, and enrollment data from September 2012 to May 2018.

Notes: The baseline period covers the 12-month period before enrollment and is beneficiary specific. Enrollment ended in February 2017.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in Medicaid or FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level, two-tailed test.

Appendix B presents full balance results. The study required an exact match on the quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, comorbid condition, and, for Medicaid beneficiaries, if they were enrolled in comprehensive managed care.

Table 3 (*continued*)

[a] Because of the high percentage of beneficiaries of unknown race among Medicaid beneficiaries, the race estimates for the Medicaid sample might not be inaccurate.

[b] To avoid double-counting, the Medicare—but not the Medicaid—analysis included all dually eligible participants.

[c] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

[d]The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms

CDPS = Chronic Illness and Disability Payment System; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end stage renal disease; HCC = hierarchical condition category; n.a.= not applicable; PBPM = per beneficiary per month.

Analytic approach

The impact estimates rely on a difference-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Assuming that external trends affect both groups similarly, a comparison group well matched on observable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables will capture differences on unobserved variables as well. The primary outcomes are total Medicare spending, number of hospital admissions, and number of ED visits. Secondary outcomes include number of primary care visits and number of specialty care visits. The awardee-specific measure was the percentage of enrollees filling a hepatitis C prescription. Regressions were estimated separately for Medicaid and Medicare beneficiaries, and pooled estimates were obtained by taking the weighted average of the Medicaid and Medicare estimates, where the weights reflect the relative size of the Medicaid/Medicare beneficiaries in the sample. Appendix A provides additional detail on the analytic approach for estimating program impacts.The study defines the pre-enrollment period as the year before each participant's enrollment date and the post-enrollment period as the following three years. It defines the enrollment date as the date on which the beneficiary signed up for the program. The study assigned a pseudo-enrollment date to each comparison beneficiary based on the enrollment date of the treatment group member to which the comparison beneficiary was matched. Appendix A describes the statistical models and outcomes used to estimate the effects of the program.

IMPACT RESULTS

Project INSPIRE had favorable estimated impacts on the awardee-specific near-term outcome. Compared to those not enrolled in Project INSPIRE, the proportion of patients with an HCV prescription fill (that is, had at least one prescription claim for a drug used to treat HCV) increased by more than **52 percentage points** (Table 4). Nearly 84 percent of enrollees had filled an HCV prescription by the end of the three-year period, compared to 30 percent of the comparison group (Appendix C, Table C.1). The high level of provider engagement likely helped Project INSPIRE affect the percentage of beneficiaries starting and completing prescription drug

treatment. According to awardee data, 76 percent of those who were candidates for treatment were adherent and 70 percent obtained an SVR.

In addition, Project INSPIRE increased the number of primary care visits by 21.0 percent for Medicare beneficiaries and by 7.7 percent for the combined sample of Medicaid and Medicare beneficiaries in the three years following beneficiary enrollment (though the increase for the combined sample was not statistically significant at the 10 percent level). There was also a 7 percent increase in specialty visits across the pooled sample, but it was not quite statistically significant. The increase in primary care and specialty visits might be due to beneficiaries visiting their doctors more frequently as part of the intervention protocols: regular appointments for HCV prescriptions and follow-up check-in and screening visits.

Main findings from impact evaluation

- Project INSPIRE increased the proportion of patients with an HCV prescription fill by more than 52 percentage points relative to comparators. Almost 84 percent of program participants filled an HCV prescription by the end of the three-year period. This is consistent with the program's primary goal of obtaining an SVR among adults with HCV via proper adherence to prescription drug treatment.
- Project INSPIRE also increased the number of primary care visits. This result could be attributable to beneficiaries visiting their doctors more frequently for regular appointments for HCV prescriptions and for follow-up check-in and screening visits—all part of the intervention protocol.

Table 4. Estimated impact of Project INSPIRE on selected outcome measures over three-year intervention period

Medicare and Medicaid			Medicaid only			Medicare only		
Impact	Percentage impact	*p*-value	Impact	Percentage impact	*p*-value	Impact	Percentage impact	*p*-value
Total Medicare expenditures ($ PBPM)								
n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-6.5	< 1%	0.98
Hospital stays, per 1,000 beneficiaries								
27	3.0%	0.75	16	2.0%	0.78	72	8.2%	0.43
ED or observation visits, per 1,000 beneficiaries								
124	3.7%	0.68	30	2.0%	0.77	138	15.0%	0.21
Percentage of beneficiaries with a readmission								
1.76	9.3%	0.61	1.7	9.8%	0.27	2	8.1%	0.53
Primary care visits in ambulatory settings, per 1,000 beneficiaries								
352	7.7%	0.20	175	4.1%	0.37	1,060**	21.0%	0.02
Specialty care visits in all settings, per 1,000 beneficiaries								
820	6.5%	0.13	790	6.3%	0.16	941	7.1%	0.14
Any hepatitis C prescriptions (percentage point impacts)								
52%***	178%	< 0.01	54%***	174%	< 0.01	53%***	177%	< 0.01

Table 4 (*continued*)

Medicare and Medicaid			Medicaid only			Medicare only		
Impact	Percentage impact	*p*-value	Impact	Percentage impact	*p*-value	Impact	Percentage impact	*p*-value
Sample size								
	Treatment 1,637	**Comparison** 6,360		**Treatment** 1,310	**Comparison** 5,085		**Treatment** 327	**Comparison** 1,275

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicare and Medicaid FFS claims, encounter, and enrollment data from September 2012 to May 2018.

Notes: Impact estimates are based on the regression-adjusted difference between the treatment and comparison group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents full impact estimates. Appendix D provides the results from the Bayesian analysis.

Expenditure data were not available because most Medicaid beneficiaries were in managed care programs.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; n.a. = not available; PBPM = per beneficiary per month.

Despite its favorable effects on HCV treatment, the program did not reduce hospitalizations, ED use, readmissions, or Medicare expenditures. Because Project INSPIRE's target population became harder to recruit in the program's second year, effects might have been less favorable for later enrollees. However, results were not sensitive to whether the enrollee was enrolled early or late in the program. They were also not sensitive to trimming outliers nor to the length of the baseline period. Thus, there is no support for a conclusion that the program reduced expenditures or service use during the observation period. Appendix C presents the full results of the impact analysis. Appendix D provides the results from the Bayesian analysis.

CONCLUSION

The results show that Project INSPIRE successfully initiated HCV drug treatment among both Medicaid and Medicare beneficiaries enrolled in the program. Adults with HCV must properly adhere to prescription drug treatment to obtain an SVR. According to the program's theory of action, care coordinators and primary care providers would encourage beneficiaries to initiate treatment and then to remain adherent to their treatment protocols. Awardee data suggest that most enrollees followed the full course of treatment, including the screening visit to confirm SVR. Enrollees also became more engaged in their treatment, as evidenced by the observed increase in the number of physician visits. Because the program expected benefits of treating HCV to occur many years after treatment, it is not surprising that the study did not observe impacts on expenditures or hospitalizations in the relatively shorter-term outcomes used here.

Limitations of evaluation

The analysis has several limitations. First, two health care systems, which could differ from other hospitals or systems in the services they offer patients with HCV, implemented the intervention.

Thus, results might not be generalizable to other settings. Second, the release of a new class of drugs to treat HCV with fewer adverse side effects coincided with better enrollment than expected and those who enrolled might have had greater motivation to initiate and adhere to HCV drug treatment. Finally, many of the benefits of treating HCV can occur many years after treatment. Although the program succeeded in improving medication adherence and most participants obtained an SVR, it is unlikely that the intervention could prevent adverse events related to HCV (and thereby reduce health care use) within the three-year evaluation window.

PROGRAM SUSTAINABILITY

The two health systems participating in DOHMH's Project INSPIRE program, Montefiore Health System and Mount Sinai Health System, sustained modified versions of the program using internal funding after the original award period ended on August 31, 2017. Although DOHMH no longer oversees the program at the two sites, the awardee reported that both health systems sustained the core elements of the program with minor modifications. The health systems sought to better tailor the program to their needs, as well as reduce costs, for example by reducing the frequency of tele-mentoring for providers from biweekly to monthly and updating the health promotion education materials to be more engaging and less time consuming.

Original payment model proposed by DOHMH

DOHMH proposed that participating sites pay for integrated care delivery services for patients with HCV using a one-time bundled payment from Medicare or Medicaid managed care organizations. The bundled payment aimed to support both the physicians' time spent in conferences with care coordinators and tele-mentoring, as well as the care coordinators' time. The awardee also proposed a separate shared savings and shared loss component to the payment model. The program assigned each site a target SVR rate and sites received bonus payments for exceeding the target or paid penalties for falling short. The awardee calculated that, based on its enrollment through February 2017, program services would cost sites an average of $760 per beneficiary.

DOHMH continued using award funding to develop its payment models until February 28, 2018, which was the end of its no-cost extension period. The awardee began pursuing ways to fund program services through FFS billing after the original payment model was not recommended to the U.S. Department of Health and Human Services Secretary when the awardee presented it at the Payment Model Technical Advisory Committee meeting in Washington, DC, on September 21, 2017. The FFS payment model uses complex care management codes to fund care coordination and tele-mentoring. DOHMH is also pursuing the development of a multipayer payment model for HCV and other comorbid conditions commonly seen in the intervention.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays rely on a difference-in-differences approach with beneficiary fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or emergency department (ED) visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or ED visit during the baseline period. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of outcomes.

The impact analysis included 59 percent of the total participants (Tables A.1 and A.2). The study defined participants as beneficiaries who enrolled in Project INSPIRE from January 2015 through February 2017, as reported in the awardee's final encounter database. The analysis dropped most of the excluded Medicare participants from the study because they were enrolled in Medicare Advantage (50 percent); the most common reason for dropping Medicaid beneficiaries was that they did not have an HCV diagnosis in the Medicaid claims (13 percent).

Table A.1. Identifying the final sample for impact analysis for FPHNY:Medicare

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total Medicare program participants through August 31, 2017		**945**
Not 18 years old (or older) on day of enrollment	0	945
Not a resident of New York City on day of enrollment	18	927
Not alive at enrollment	5	922
Lack of Medicare enrollment (Part A and B) on HCIA program enrollment date	69	853
Enrolled in Medicare Advantage	476	377
Medicare was not primary payer on day of enrollment	8	369
Did not have at least 90 days of Medicare FFS enrollment (Part A and B) in the baseline period	32	337
Did not have hepatitis C diagnosis on or before day of enrollment	9	328
Not alive 30 days after enrollment date	1	327
Final Medicare analytic sample		**327**

Sources: Mathematica's analysis of information from the awardee's program encounter database from January 2015 through February 2017, and Medicare claims and enrollment data from September 2012 through May 2018, as of March 13, 2019.

Notes: There were 681 Medicaid beneficiaries that were dually eligible for Medicare. They are included in the Medicare sample.

HCIA = Health Care Innovation Awards; FPHNY = Fund for Public Health in New York; FFS = fee-for-service.

Table A.2. Identifying the final sample for impact analysis for FPHNY: Medicaid

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total Medicaid non-dual program participants through August 31, 2017		**1,830**
Did not have hepatitis C diagnosis on or before day of enrollment	242	1,588
Not 18 or older or not a resident of New York City	170	1,418
Beneficiaries who died, had private insurance, restricted benefits, enrolled in State Children's Health Insurance Program	90	1,357
Beneficiaries who lacked 90 days of Medicaid enrollment during baseline period	14	1,343
Not alive 30 days after enrollment date	2	1,312
Final Medicaid analytic sample		**1,310**

Sources: Mathematica's analysis of information from the awardee's program encounter database from January 2015 through February 2017, and Medicaid claims and enrollment data from September 2012 through May 2018, as of March 13, 2019.

Notes: There were 681 Medicaid beneficiaries that were dually eligible for Medicare. They are included in the Medicare sample.

HCIA = Health Care Innovation Awards; FPHNY = Fund for Public Health in New York; FFS = fee-for-service.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Tables B.1 and B.2 show the variables used for matching the Medicaid and Medicare samples, respectively. Table B.1 displays the weighted means of baseline characteristics for the 1,310 treatment beneficiaries and the 5,085 matched comparison beneficiaries used in the Medicaid impact analysis; Table B.2 displays the weighted means of baseline characteristics for the 327 treatment beneficiaries and the 1,275 matched comparison beneficiaries used in the Medicare impact analysis. These tables show the means, difference in means, the percentage difference, and the standardized difference for each variable, which the study calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables include demographic characteristics (age, gender, and race); Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category [HCC] score); aggregate Medicare expenditures; and service use. The study required an exact match on the quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, comorbid condition, and, for Medicaid beneficiaries, if they were enrolled in comprehensive managed care. The analysis measured variables over various specified intervals within the 12 months before enrollment in the intervention. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

Table B.1 shows the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence of the statistical significance of the difference in the means. The equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the study also performed an omnibus test in which the null hypothesis is that the treatment and matched comparison groups are balanced across all linear combinations of the covariates. The results assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

Table B.1. Baseline characteristics of treatment and matched comparison groups for FPHNY: Medicare sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	64 (0.50)	65 (0.30)	-0.38 (0.78)	< +/-1	-0.04	0.63	< 0.01
Female, %	32 (2.6)	35 (1.3)	-3.0 (3.7)	-2.3	-0.06	0.42	< 0.01
White, %	32 (2.6)	33 (1.3)	-0.80 (3.6)	-2.5	-0.02	0.83	< 0.01
Black, %	49 (2.8)	49 (1.4)	-0.07 (3.8)	< +/-1	0.00	0.99	< 0.01
Hispanic, %	14 (1.9)	13 (0.91)	1.1 (2.6)	7.6	0.03	0.68	< 0.01
Medicare entitlement and dual eligibility status, %							
Dually eligible for Medicare and Medicaid	75 (2.4)	74 (1.3)	0.89 (3.4)	1.2	0.02	0.80	< 0.01
Original reason for Medicare entitlement: age	43 (2.7)	45 (1.4)	-2.3 (3.8)	-5.5	-0.05	0.54	< 0.01
Original reason for Medicare entitlement: disability	51 (2.8)	49 (1.4)	2.6 (3.7)	5.1	0.05	0.48	< 0.01
Part D coverage in the month before enrollment	91 (1.6)	91 (0.79)	-0.50 (2.2)	< +/-1	-0.02	0.82	< 0.01
Health status and diagnosis							
HCC score[a]	2.19 (0.09)	2.07 (0.04)	0.11 (0.12)	5.2	0.07	0.34	< 0.01
Had cirrhosis, %	29 (2.5)	27 (1.2)	2.1 (3.6)	7.2	0.05	0.56	< 0.01
Had hepatitis C prescription during 2 years before enrollment, %	18 (2.1)	16 (1.0)	2.1 (3.0)	12	0.05	0.49	< 0.01
Had hepatocellular carcinoma, %	4.0 (1.1)	3.1 (0.50)	0.92 (1.4)	23	0.05	0.53	< 0.01
Had liver transplant, %	2.1 (0.80)	2.1 (0.40)	0.07 (1.1)	3.1	0.00	0.95	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Medicare expenditures							
Total expenditures[b] (year before enrollment, annualized)	2,767 (228)	2,658 (109)	109 (321)	3.9	0.03	0.73	< 0.01
Total expenditures, 1st baseline year[b] (2 years before enrollment, annualized)	2,123 (n.a.)	2,017 (n.a.)	106 (315)	5.0	0.03	0.68	< 0.01
Total expenditures, 3 months before enrollment[b] (annualized)	3,217 (357)	3,108 (158)	110 (494)	3.4	0.02	0.82	< 0.01
Service utilization							
1 hospitalization, year before enrollment, %	22 (2.3)	21 (1.1)	0.95 (3.3)	4.4	0.02	0.77	< 0.01
2 hospitalizations, year before enrollment, %	6.7 (1.4)	6.6 (0.69)	0.17 (2.0)	2.5	0.01	0.93	< 0.01
3 or more hospitalizations, year before enrollment, %	11 (1.8)	10 (0.81)	1.1 (2.4)	10	0.04	0.64	< 0.01
1 hospitalization, 1st baseline year, %	13 (1.9)	12 (0.88)	0.94 (2.7)	7.3	0.03	0.72	< 0.01
2 hospitalizations, 1st baseline year, %	8.9 (1.6)	9.1 (0.76)	-0.25 (2.3)	-2.9	-0.01	0.91	< 0.01
3 or more hospitalizations, 1st baseline year, %	8.0 (1.5)	6.3 (0.66)	1.7 (2.0)	21	0.06	0.41	< 0.01
1 hospitalization, 3 months before enrollment, %	13 (1.9)	14 (0.95)	-0.71 (2.7)	-5.4	-0.02	0.79	< 0.01
2 hospitalizations, 3 months before enrollment, %	4.9 (1.2)	4.3 (0.53)	0.63 (1.6)	13	0.03	0.70	< 0.01
3 or more hospitalizations, 3 months before enrollment, %	1.5 (0.68)	1.3 (0.30)	0.28 (0.97)	18	0.02	0.78	< 0.01
Had any hospitalization for mental health or substance abuse, %	6.4 (1.4)	4.8 (0.56)	1.7 (1.8)	26	0.07	0.37	< 0.01
Total ED or observation visits during year before enrollment[b] (annualized; per 1,000 beneficiaries)	912 (86)	949 (42)	-37 (121)	-4.1	-0.02	0.76	< 0.01
Total ED or observation visits during 2 years before enrollment[b] (annualized; per 1,000 beneficiaries)	770 (n.a.)	668 (n.a.)	103 (146)	13	0.06	0.38	0.02

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
ED or observation visits, had any visit 30 days before enrollment, %	4.3 (1.1)	4.9 (0.59)	-0.62 (1.7)	-14	-0.03	0.71	< 0.01
ED or observation visits, had any visit 31 to 90 days before enrollment, %	15 (2.0)	17 (1.00)	-2.1 (3.0)	-14	-0.06	0.48	0.01
Had any ED or observation visit for mental health or substance abuse, %	4.9 (1.2)	4.0 (0.54)	0.88 (1.6)	18	0.04	0.59	< 0.01
Primary care visits,[b] any setting (annualized; per 1,000 beneficiaries)	7,371 (499)	7,328 (234)	43 (703)	< +/-1	0.00	0.95	< 0.01
Primary care visits,[b] any setting, 30 days before enrollment (annualized; per 1,000 beneficiaries)	10,171 (918)	10,081 (435)	90 (1,273)	< +/-1	0.01	0.94	< 0.01
Primary care visits,[b] any setting, 31 to 90 days before enrollment. (annualized; per 1,000 beneficiaries)	6,949 (594)	7,122 (286)	-174 (862)	-2.5	-0.02	0.84	< 0.01
Specialist visits,[b] any setting (annualized; per 1,000 beneficiaries)	17,422 (886)	17,182 (419)	239 (1,199)	1.4	0.01	0.84	< 0.01
Specialist visits,[b] any setting, 30 days before enrollment (annualized; per 1,000 beneficiaries)	25,040 (1,779)	25,234 (891)	-193 (2,448)	< +/-1	-0.01	0.94	< 0.01
Specialist visits,[b] any setting, 31 to 90 days before enrollment, (annualized; per 1,000 beneficiaries)	17,440 (1,174)	17,458 (563)	-18 (1,646)	< +/-1	0.00	0.99	< 0.01
Had any visit with a hepatitis C specialist, %	70 (2.5)	71 (1.3)	-1.4 (3.4)	-2.0	-0.03	0.68	< 0.01
Had any inpatient visit 30 days before enrollment, %	8.3 (1.5)	7.7 (0.73)	0.55 (2.1)	6.6	0.02	0.79	< 0.01
Had any outpatient visit 30 days before enrollment, %	83 (2.1)	87 (1.00)	-3.6 (2.8)	-4.3	-0.10	0.20	0.02
Had any physician visit, 30 days before enrollment, %	84 (2.0)	85 (1.0)	-1.7 (2.8)	-2.0	-0.05	0.55	< 0.01
Had any inpatient, outpatient, or physician visit, 30 days before enrollment, %	92 (1.5)	94 (0.71)	-1.8 (2.0)	-2.0	-0.07	0.36	< 0.01
Saw same physician more than once during year before enrollment, %	98 (0.86)	99 (0.36)	-1.1 (1.1)	-1.1	-0.07	0.31	< 0.01
Propensity score	-363.67 (7.34)	-367.69 (3.59)	4.03 (10.37)	-1.1	0.03	0.70	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Number of beneficiaries	327	1,275					
Omnibus test				Chi-squared statistic 18.66	Degrees of freedom 44.00	P-value 1.00	

Source: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS claims from September 2012 to May 2018.

Note : Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include each enrollment quarter and comorbid condition.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

ED = emergency department; FPHNY = Fund for Public Health in New York; HCC = hierarchical condition category; SE = standard error.

Table B.2. Baseline characteristics of treatment and matched comparison groups for FPHNY: Medicaid sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	52 (0.24)	53 (0.14)	-0.23 (0.37)	< +/-1	-0.03	0.54	< 0.01
Male, %	65 (1.3)	65 (0.67)	-0.36 (2.0)	< +/-1	-0.01	0.85	< 0.01
Hispanic, %	48 (1.4)	45 (0.69)	3.0 (1.9)	6.2	0.06	0.12	< 0.01
Medicaid enrollment, %							
Medicaid enrolled for 365 days during year before enrollment	90 (0.83)	89 (0.44)	0.98 (1.2)	1.1	0.03	0.41	< 0.01
Medicaid enrolled for 365 days during pre-baseline year (two years before enrollment)	79 (1.1)	77 (0.59)	1.3 (1.6)	1.7	0.03	0.41	< 0.01
Health status and diagnosis							
CDPS score[a]	4.5 (0.07)	4.5 (0.03)	-0.04 (0.10)	< +/-1	-0.01	0.71	< 0.01
AIDS or other infectious disease, %	97 (0.49)	98 (0.24)	-0.85 (0.63)	< +/-1	-0.05	0.18	< 0.01
Cardiovascular disease, %	58 (1.4)	59 (0.69)	-0.82 (2.0)	-1.4	-0.02	0.68	< 0.01
Cirrhosis, %	19 (1.1)	17 (0.39)	1.4 (1.5)	7.4	0.04	0.37	< 0.01
Diabetes, %	27 (1.2)	29 (0.63)	-1.7 (1.7)	-6.5	-0.04	0.32	< 0.01
Disabled, %	48 (1.4)	47 (0.70)	1.3 (2.0)	2.6	0.03	0.53	< 0.01
Hepatocellular carcinoma, %	1.6 (0.35)	1.5 (0.12)	0.13 (0.45)	7.9	0.01	0.78	< 0.01
Liver transplant, %	0.69 (0.23)	0.67 (0.08)	0.02 (0.31)	2.2	0.00	0.96	< 0.01
Psychiatric condition, %	47 (1.4)	47 (0.70)	0.03 (1.9)	< +/-1	0.00	0.99	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Substance abuse, %	67 (1.3)	67 (0.67)	0.31 (1.9)	< +/-1	0.01	0.87	< 0.01
Any comorbidity in baseline year, %	92 (0.76)	92 (0.36)	-0.15 (1.1)	< +/-1	-0.01	0.89	< 0.01
Service use							
0 hospitalizations, year before enrollment, %	67 (1.3)	67 (0.63)	-0.56 (1.8)	< +/-1	-0.01	0.75	< 0.01
1 hospitalization, year before enrollment, %	18 (1.1)	17 (0.52)	0.67 (1.4)	3.8	0.02	0.64	< 0.01
2 hospitalizations, year before enrollment, %	6.9 (0.70)	7.5 (0.33)	-0.56 (0.99)	-8.0	-0.02	0.57	< 0.01
3 or more hospitalizations, year before enrollment, %	8.4 (0.77)	7.9 (0.33)	0.44 (1.1)	5.3	0.02	0.68	< 0.01
0 hospitalizations, 3 months before enrollment, %	86 (0.95)	86 (0.43)	-0.14 (1.3)	< +/-1	0.00	0.92	< 0.01
1 hospitalization, 3 months before enrollment, %	11 (0.85)	11 (0.39)	0.14 (1.2)	1.3	0.00	0.91	< 0.01
2 hospitalizations, 3 months before enrollment, %	1.9 (0.38)	1.6 (0.17)	0.30 (0.50)	16	0.02	0.55	< 0.01
3 or more hospitalizations, 3 months before enrollment, %	1.1 (0.29)	1.4 (0.13)	-0.30 (0.44)	-26	-0.03	0.50	< 0.01
Any hospitalization, 30 days before enrollment, %	5.6 (0.64)	5.5 (0.28)	0.10 (0.91)	1.8	0.00	0.91	< 0.01
Any hospitalization for mental health or substance abuse, %	9.8 (0.82)	9.7 (0.41)	0.06 (1.2)	< +/-1	0.00	0.96	< 0.01
Total ED or observation visits[a] (annualized; per beneficiary)	1.2 (0.05)	1.2 (0.02)	0.04 (0.07)	3.2	0.02	0.61	< 0.01
Had any ED or observation visit, 30 days before enrollment, %	9.1 (0.79)	9.2 (0.38)	-0.08 (1.1)	< +/-1	0.00	0.94	< 0.01
Had at least one ED or observation visit, 31 to 90 days before enrollment, %	15 (0.98)	15 (0.48)	-0.09 (1.4)	< +/-1	0.00	0.95	< 0.01
Had any ED or observation visit for mental health or substance abuse during year before enrollment, %	8.5 (0.77)	7.9 (0.37)	0.63 (1.1)	7.4	0.02	0.56	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Primary care visits in any setting during year before enrollment[b] (per beneficiary)	6.3 (0.22)	6.1 (0.10)	0.25 (0.29)	3.9	0.03	0.40	< 0.01
Primary care visits in any setting, 30 days before enrollment[b] (annualized; per beneficiary)	0.63 (0.03)	0.63 (0.01)	0.00 (0.04)	< +/-1	0.00	0.98	< 0.01
Primary care visits, any setting, 31 to 90 days before enrollment[b] (annualized; per beneficiary)	1.0 (0.05)	0.97 (0.02)	0.04 (0.06)	3.9	0.02	0.54	< 0.01
Specialist visits, any setting[b] (annualized; per beneficiary)	10 (0.34)	10 (0.16)	0.05 (0.47)	< +/-1	0.00	0.92	< 0.01
Specialist visits in any setting, 30 days before enrollment[b] (annualized; per beneficiary)	1.1 (0.05)	1.1 (0.02)	-0.04 (0.07)	-4.2	-0.03	0.51	< 0.01
Specialist visits in any setting, 31 to 90 days before enrollment[b] (annualized; per beneficiary)	1.7 (0.08)	1.7 (0.04)	0.00 (0.11)	< +/-1	0.00	0.98	< 0.01
Had any visit with a hepatitis C specialist, %	40 (1.4)	40 (0.67)	0.29 (2.0)	< +/-1	0.01	0.88	< 0.01
Had hepatitis C prescription, 2 years before enrollment, %	12 (0.91)	11 (0.45)	1.3 (1.3)	11	0.04	0.29	< 0.01
Had any 30-day unplanned readmission, %	7.5 (0.73)	6.8 (0.31)	0.77 (1.0)	10	0.03	0.45	< 0.01
Had any physician visit, 30 days before enrollment, %	64 (1.3)	65 (0.68)	-0.57 (1.9)	< +/-1	-0.01	0.76	< 0.01
Had 2 or more visits with the same provider during year before enrollment, %	82 (1.1)	81 (0.55)	0.81 (1.6)	< +/-1	0.02	0.60	< 0.01
Had any inpatient, outpatient, or physician visit, 30 days before enrollment, %	65 (1.3)	66 (0.68)	-0.45 (1.8)	< +/-1	-0.01	0.81	< 0.01
Propensity score	0.04 (0.00)	0.04 (0.00)	0.00 (0.00)	4.7	0.04	0.34	< 0.01
Number of beneficiaries	1,312	5,094					
Omnibus test				Chi-squared statistic 655.45	Degrees of freedom 116.00	P-value 0.00	

Source: Mathematica's analysis of awardee-provided enrollment data and Medicaid FFS claims, encounter, and enrollment data from September 2012 to May 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Table B.2 (*continued*)

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid. Exact matching variables include enrollment in comprehensive managed care and quarter of enrollment.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending

[b]Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; FFS = fee-for-service; FPHNY = Fund for Public Health in New York; HCC = hierarchical condition category; SE = standard error.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Tables C.1 and C.2 display the results from the impact analysis. Regressions were estimated separately for Medicaid and Medicare beneficiaries, and pooled estimates were obtained by taking the weighted average of the Medicaid and Medicare estimates, where the weights reflect the relative size of the Medicaid/Medicare beneficiaries in the sample. Table C.1 shows the impact estimates for the combined sample of Medicaid and Medicare beneficiaries. Table C.2 shows the impact estimates for Medicaid and Medicare beneficiaries, measured separately over intervention Years 1 through 3. The analysis estimated models over Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the .01, .05, and .10 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the FPHNY intervention on selected use measures during one-, two-, and three-year follow-up periods

	Medicare and Medicaid, combined sample				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries					
Baseline year	828	718			
Year 1	897	750	36.0 (73)	4.2%	0.62
Year 2	946	880	-43.5 (86)	-4.4%	0.61
Year 3	1,136	902	126.1 (129)	12%	0.33
Cumulative	924	787	27.2 (65)	3.0%	0.67
ED or observation visits, per 1,000 beneficiaries					
Baseline year	1,294	1,117			
Year 1	1,407	1,279	71 (138)	5.3%	0.61
Year 2	1,514	1,396	60 (155)	4.1%	0.70
Year 3	1,551	1,504	12 (257)	-0.8%	0.96
Cumulative	1,451	1,342	52 (124)	3.7%	0.68
Percentage of beneficiaries with any hospital admission in a time period					
Baseline year	35	34			
Year 1	34	30	3.5 (2.0)	11.4%	0.07
Year 2	33	31	2.1 (2.0)	6.8%	0.35
Year 3	33	29	3.5 (4.0)	12.0%	0.33
Cumulative	58	55	3.6 (2.0)	6.6%	0.09
Percentage of beneficiaries with any ED or observation visits in a time period					
Baseline year	47	47			
Year 1	47	45	2.0 (2.0)	4.4%	0.36
Year 2	46	44	2.3 (2.0)	5.3%	0.33
Year 3	45	45	-0.6 (4.0)	-1.3%	0.87
Cumulative	72	70	2.5 (2.0)	3.5%	0.21
Primary care visits in ambulatory setting, per 1,000 beneficiaries					
Baseline year	4,969	4,903			
Year 1	5,159	4,575	518.0* (269)	11.2%	0.05
Year 2	4,861	4,378	425.1 (391)	9.6%	0.28

Table C.1 (*continued*)

	Medicare and Medicaid, combined sample				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Year 3	4,261	4,462	-230.1 (504)	-5.1%	0.65
Cumulative	4,902	4,483	352 (271)	7.7%	0.19
Specialist visits in all settings, per 1,000 beneficiaries					
Baseline year	11,549	11,578			
Year 1	13,438	12,444	1,022 (561)	8.2%	0.07
Year 2	13,433	13,054	408 (698)	3.1%	0.56
Year 3	14,314	13,089	1,258 (1,111)	9.6%	0.26
Cumulative	13,408	12,615	820 (548)	6.5%	0.13
Any hepatitis C prescriptions (percentage point impacts)					
Baseline year	11	9			
Year 1	68	14	53.8*** (2.0)	376%	0.00
Year 2	15	9	5.9** (2.0)	66%	0.00
Year 3	9	9	0.1 (2.0)	1.0%	0.97
Cumulative	84	30	53.8*** (2.0)	179%	0.00
Percentage of beneficiaries with a hospital readmission					
Baseline year	9	7			
Year 1	9	8	1.2 (1.0)	15%	0.33
Year 2	10	8	1.2 (1.0)	15%	0.41
Year 3	10	8	2.0 (2.0)	26%	0.41
Cumulative	21	19	1.8 (2.0)	9.3%	0.38
Sample sizes					
Number of beneficiaries					
Year 1	1,637	6,360			
Year 2	1,293	4,762			
Year 3	621	2,248			
Cumulative	1,637	6,360			

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicare and Medicaid FFS claims, encounter, and enrollment data from September 2012 to May 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiaries' characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FPHNY = Fund for Public Health in New York; PBPM = per beneficiary per month; SE = standard error.

Table C.2. Estimated impact of the FPHNY intervention on selected Medicare FFS expenditures (dollars PBPM) and Medicare and Medicaid use measures during one-, two-, and three-year follow-up periods

	Medicare					Medicaid				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)										
Baseline year	2,933	2,756								
Year 1	3,524	3,364	-17 (321)	< 1%	0.96					
Year 2	3,565	3,715	-327 (396)	-9.8%	0.41					
Year 3	4,012	3,608	227 (541)	8.3%	0.67					
Cumulative	3,478	3,307	-6.5 (277)	< 1%	0.98					
Hospital stays, per 1,000 beneficiaries										
Baseline year	961	824				795	691			
Year 1	1,139	910	92 (101)	9.3%	0.36	836	710	22 (64)	2.8%	0.73
Year 2	1,100	1,016	-53 (129)	-6.0%	0.68	906	844	-41 (70)	-5.1%	0.55
Year 3	1,343	996	210 (178)	35%	0.24	1,079	874	102 (111)	14%	0.36
Cumulative	1,114	905	72 (91)	8.2%	0.43	877	758	16 (56)	2.0%	0.78
ED or observation visits, per 1,000 beneficiaries										
Baseline year	1,069	967				1,350	1,304			
Year 1	1,291	930	258* (133)	26%	0.05	1,436	1,366	24 (107)	1.6%	0.82
Year 2	971	933	-65 (132)	-7.6%	0.62	1,657	1,518	93 (134)	6.3%	0.49
Year 3	1,139	1,022	15 (240)	1.4%	0.95	1,670	1,644	-20 (241)	-1.4%	0.93
Cumulative	1,165	925	138 (110)	15%	0.21	1,523	1,447	30 (102)	2.0%	0.77
Percentage of beneficiaries with any hospital admission in a time period										
Baseline year	41	39				34	33			
Year 1	39	36	2.3 (3.1)	6.4%	0.44	33	29	3.8** (1.5)	13%	0.01
Year 2	38	34	3.3 (3.5)	9.4%	0.35	32	30	1.8 (1.8)	6.0%	0.32
Year 3	36	30	6 (5.4)	20%	0.27	32	29	2.8 (2.9)	9.5%	0.34
Cumulative	63	58	4.1 (3.3)	7.1%	0.21	57	54	3.5** (1.7)	6.6%	0.04

Table C.2 (*continued*)

	Medicare					Medicaid				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Percentage of beneficiaries with any ED or observation visits in a time period										
Baseline year	44	46				48	47			
Year 1	45	39	5.5 (3.4)	14%	0.11	47	46	1.1 (1.7)	2.4%	0.51
Year 2	40	39	0.57 (3.8)	1.5%	0.88	48	45	2.8 (1.9)	6.3%	0.14
Year 3	45	38	6.4 (5.4)	17%	0.24	45	47	-2.6 (2.9)	-5.5%	0.37
Cumulative	70	67	3.1 (3.2)	4.6%	0.34	73	71	2.3 (1.5)	3.2%	0.13
Primary care visits in ambulatory setting, per 1,000 beneficiaries										
Baseline year	5,083	5,473				4,941	4,760			
Year 1	6,257	5,621	1025** (459)	20%	0.03	4,885	4,313	391** (195)	8.7%	0.04
Year 2	6,671	5,625	1436** (677)	29%	0.03	4,392	4,050	160 (270)	3.9%	0.55
Year 3	6,331	5,918	802 (859)	17%	0.35	3,692	4,038	-527 (338)	-13%	0.12
Cumulative	6,273	5,602	1060** (462)	21%	0.02	4,560	4,203	175 (196)	4.1%	0.37
Specialist visits in all settings, per 1,000 beneficiaries										
Baseline year	17,700	17,652				11,274	11,254			
Year 1	22,211	20,929	1,235 (1,235)	6.0%	0.32	13,147	12,141	985* (578)	8.1%	0.09
Year 2	22,314	21,309	958 (1,625)	5.0%	0.56	13,439	13,161	257 (735)	2.0%	0.73
Year 3	23,935	21,456	2,432 (2,141)	13%	0.26	14,688	13,145	1,522 (1,235)	12%	0.22
Cumulative	21,998	20,484	1,466 (1,105)	7.5%	0.18	13,279	12,468	790 (568)	6.3%	0.16
Any hepatitis C prescriptions (percentage point impacts)										
Baseline year	17	13				10	8			
Year 1	69	16	53*** (2.9)	331%	0.001	68	14	54*** (1.5)	385%	0.001
Year 2	13	9	4.7* (2.5)	57%	0.06	15	9	6.2*** (1.5)	70%	0.001
Year 3	11	9	1.9 (3.4)	21%	0.57	9	9	-0.43 (2.1)	-4.8%	0.83
Cumulative	83	31	53*** (2.8)	177%	0.001	84	30	54*** (1.4)	180%	0.001

Table C.2 (*continued*)

	Medicare					Medicaid				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Percentage of beneficiaries with a hospital readmission										
Baseline year	12	9				8	7			
Year 1	15	12	2.3 (2.1)	18%	0.28	8	7	0.95 (0.93)	13%	0.31
Year 2	10	9	0.93 (2.3)	11%	0.68	10	8	1.3 (1.2)	16%	0.28
Year 3	10	8	1.3 (3.4)	16%	0.7	10	8	2.2 (2.1)	29%	0.29
Cumulative	27	24	2 (3.2)	8.1%	0.53	19	18	1.7 (1.6)	9.8%	0.27
Sample sizes										
Number of beneficiaries										
Year 1	327	1,275				1,310	5,085			
Year 2	266	992				1,027	3,770			
Year 3	134	507				487	1,741			
Cumulative	327	1,275				1,310	5,085			

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicare and Medicaid FFS claims, encounter, and enrollment data from September 2012 to May 2018.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for beneficiaries' characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; FPHNY = Fund for Public Health in New York; PBPM = per beneficiary per month; SE = standard error.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for the Fund for Public Health in New York (FPHNY) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to FPHNY. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on CMS's four core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for CMS's four core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for FPHNY led to a Bayesian estimate of the program's impact on total Medicare expenditures of 1 percent (an estimated increase of $43 per beneficiary per month) across the first three years of the program.

Table D.1. Comparison of frequentist and Bayesian impact estimates for FPHNY in the first three years after enrollment

		Impact estimate (95 percent interval)		Percentage impacts		
Payer	Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Medicaid	Hospital admissions	16 (-94, 126)	12 (-51, 77)	> -1%	2%	1%
	ED visits	30 (-169, 229)	15 (-91, 127)	-1%	2%	1%
	Readmissions	1.7 (-1.4, 4.8)	0.3 (-1.4, 2.2)	> -1%	10%	1%
Medicare	Total expenditures ($ PBPM)	-6.5 (-549, 536)	43 (-201, 313)	-1%	> -1%	1%
	Hospital admissions	72 (-107, 251)	15 (-59, 94)	> -1%	8%	1%
	ED visits	138 (-77, 353)	11 (-61, 91)	-1%	15%	1%
	Readmissions	2.0 (-4.3, 8.3)	0.2 (-1.1, 1.5)	> -1%	8%	1%
Pooled	Total expenditures ($ PBPM)	NA	NA	NA	NA	NA
	Hospital admissions	27 (-100, 155)	13 (-43, 70)	> -1%	3%	1%
	ED visits	52 (-152, 255)	14 (-74, 105)	-1%	4%	1%
	Readmissions	1.8 (-2.1, 5.7)	0.3 (-0.9, 1.4)	> -1%	9%	1%

Source: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS claims from September 2012 to May 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending and are evaluated for the Medicare sample only. Readmissions are the percentage of beneficiaries with a

Table D.1 (*continued*)

readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

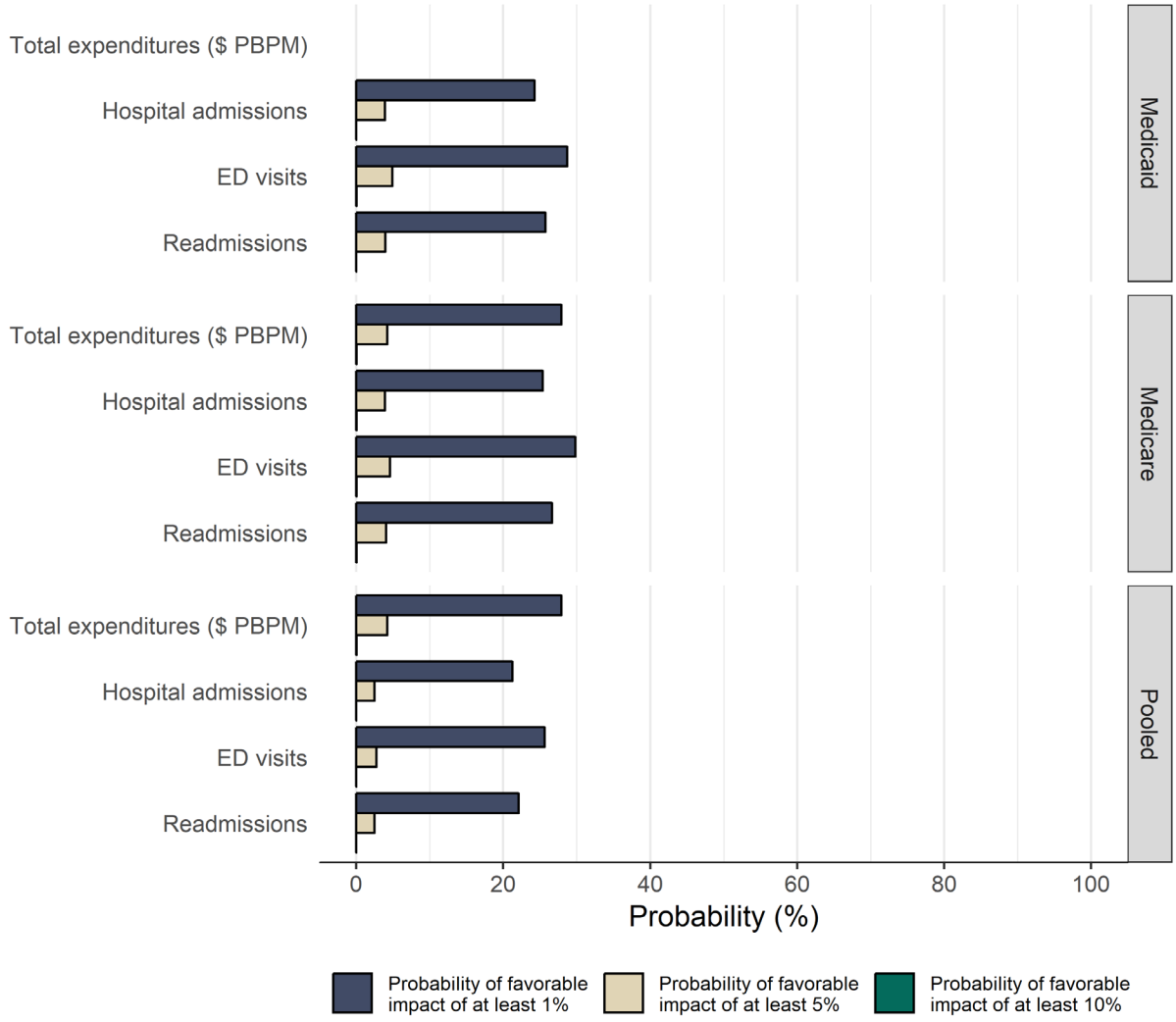
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month. NA=not available.

Because the frequentist results are imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates that are more consistent across outcomes. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that FPHNY achieved favorable impacts across the three-year follow-up period on three core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the FPHNY program had a favorable impact on key outcomes



Source: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS claims from September 2012 to May 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending and are evaluated for the Medicare sample only. Readmissions are the percentage of beneficiaries with a readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a small probability—in the range of 20-30 percent—that FPHNY had a favorable impact of 1 percent or more on each of CMS's four core outcomes, with similar probabilities for the Medicare, Medicaid, and pooled samples. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the FPHNY program did not have a meaningful impact on total expenditures or service use over the three-year study period.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Icahn School of Medicine at Mount Sinai

September 2020

Jia Pu, Kristin Geonnotti, Suzie Rastgoufard, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

ICAHN SCHOOL OF MEDICINE AT MOUNT SINAI

The Icahn School of Medicine at the Mount Sinai Hospital received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Hospital at Home (HaH) program and develop an associated payment model. The program provided acute care services in patients' homes. The target population consisted of adults who presented at a Mount Sinai Hospital emergency department (ED) or outpatient setting, lived at home in Manhattan, met the Milliman Care Guidelines (MCG) admission criteria for their conditions, and could be safely cared for at home. Mount Sinai Hospital launched the HaH program in November 2014. The intervention period funded by HCIA R2 ended in August 2017. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The HaH program aimed to improve patients' care, reduce the rate of complications, and lower the cost of care by providing acute and post-acute care services to patients in the home.
- The awardee expanded the HaH program to provide palliative and subacute rehabilitation services in the home and care for patients who were averse to entering the hospital. Given the small number of participants and late implementation of these additional services, the analysis focused on beneficiaries who received acute and post-acute care services of the program.
- Given the small sample size and likelihood of selection bias, this study does not present estimates of program impacts on outcomes.

The awardee hypothesized that acute care services provided in the home by the Mobile Acute Care Team (MACT) would produce health outcomes as good or better than those the participants would have experienced had they been hospitalized. The alternative model for traditional acute care—including both the acute care services that patients would have otherwise received in the hospital plus any post-acute services deemed appropriate by MACT staff—aimed to lower costs, improve process and clinical health outcomes, and increase participants' satisfaction.

The awardee developed a bundled payment model with risk sharing. The payment model covered program services for patients receiving acute, palliative, and observation care. The payment covered MACT services in the acute and post-acute phases, including the MACT doctors' professional fees, nursing care, social work services, and community paramedicine visits. Services not covered in the bundle (such as inpatient consultations and post-acute radiology services) were billed separately under fee-for-service (FFS) Medicare. The HaH provider was eligible for shared savings or responsible for repayment of a percentage of the annual difference between the total spending in HaH episodes and the benchmark.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The HaH program provided acute and post-acute care services in a patient's home. Through the three-year cooperative agreement, Mount Sinai extensively modified the program to increase enrollment and enhance care provided to participants. By the end of the cooperative agreement, the HaH program evolved to also provide palliative and subacute rehabilitation care and care for patients averse to entering the hospital.
Major innovation	The program provided acute care services in the home.
Program components	• Delivery of acute care services in the home • Delivery of post-acute care services in the home
Target population	The program sought to engage adults enrolled in Medicare FFS, Healthfirst Medicare, Healthfirst Medicaid managed care, or the Healthfirst health maintenance organization; presented at the Mount Sinai Hospitals, Mount Sinai St Luke's Hospital EDs, or its outpatient settings; lived at home in Manhattan, met the MCG admission criteria for their conditions; and could be safely cared for at home.
Participating providers	Mount Sinai Hospital and Mount Sinai St Luke's Hospital EDs
Total enrollment	The HaH program enrolled 295 patients from November 2014 through August 2017. Mount Sinai had enrolled only a little more than half of its target by the end of the cooperative agreement.
Theory of change or theory of action	The awardee intended the provision of acute and post-acute care services in the home to increase participants' satisfaction and result in lower costs, superior processes, and better care experience and clinical health outcomes.
Award amount	$9,610,517
Effective launch date	The program began operating in November 2014.
Program settings	Participants' homes
Market area	Manhattan, New York
Target outcomes	Clinical outcomes • Decrease complications of care (for example, reduce the rate of falls) • Shorten the length of stay • Improve care process measures • Decrease the mortality rate Cost and resource use • Decrease 30-day unplanned readmissions • Decrease total Medicare Part A and B payments • Decrease rate of hospital ED visits • Decrease rate of admissions to SNFs Care experience • Increase patient satisfaction
Payment model	Mount Sinai developed a bundled payment model with risk sharing that it could modify to meet the needs and preferences of individual payers.

Table 1 (*continued*)

Program characteristics	Description
Sustainability plans	Mount Sinai stopped providing HaH services to Medicare beneficiaries at the end of the award due to lack of federal funding. However, the hospital continued to provide HaH services to eligible patients covered by two commercial payers.

ED = emergency department; FFS = fee-for-service; HaH = Hospital at Home program; MACT = Mobile Acute Care Team; MCG = Milliman Care Guidelines; SNF = skilled nursing facility.

A rigorous impact evaluation of the HaH program was not possible because of the inability to replicate the enrollment criteria in claims data and the small number of enrollees. Therefore, this report describes only the demographic and health characteristics of Medicare FFS participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Key features of the descriptive analysis

Program characteristics	Description
Descriptive analysis	The evaluation could not replicate eligibility criteria—one of which required clinicians' judgment and another required a home assessment—using Medicare claims data or clinical registry data. Thus, it was not possible to identify a comparison group that was similar to the intervention group at the time of enrollment. This limitation, together with the small sample size, made it impossible to conduct a rigorous impact evaluation of this program. As a result, this report describes only the demographic and health characteristics of Medicare FFS participants before they enrolled in the program.
Intervention group for descriptive analysis	The intervention group for the evaluation relied on the 184 participants (among the total enrollment of 295) who were enrolled in Medicare FFS for at least three months before and after enrolling in the program. The 111 excluded beneficiaries included 50 patients who were not Medicare beneficiaries and 61 Medicare beneficiaries who were not enrolled in Medicare FFS for at least three months before and after enrollment.
Limitations	Due to the problems noted above, this report cannot make inferences about the impact of this program on Medicare costs or other program outcomes.

FFS = fee-for-service.

PROGRAM DESIGN AND ADAPTATION

The HaH program had two key phases: an acute care phase and a post-acute care phase. During each phase, the patient was eligible to receive a range of intervention services.[1]

Acute care phase

During the acute care phase, the MACT team provided acute care services to patients in their homes, including a physical examination, monitoring of illness and vital signs, intravenous infusions, wound care, and education regarding the patient's illness. If needed, the team also

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

provided durable medical equipment (DME), phlebotomy, and home radiography. During the acute care phase, nurses provided home visits once or twice daily. A physician or nurse practitioner also saw patients daily either in person or via video call. In addition, a social worker visited patients to provide services to meet their social needs. The program matched the frequency, intensity, and duration of MACT services provided with the patient's clinical needs.

Post-acute care phase

After patients were discharged from the acute care phase, the team provided transitional and follow-up services during a 30-day post-acute care phase to promote recovery from the acute episode, prevent hospital readmission, and transition to primary care. These services included radiology, lab work, nursing care, DME, pharmacy and infusion services, and telemedicine. All participants received necessary transitional services, such as physician and nurse visits, outpatient care coordination, and social work services, to prevent hospital readmissions. In addition, the program encouraged patients to reach out to the MACT team if their health status changed or medical problems arose during this time. A clinician would triage the issue over the phone or send a nurse or physician to visit the patient, if necessary. The program also encouraged patients to follow up with their regular primary care providers to receive nonacute medical care.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Mount Sinai faced challenges meeting its initial enrollment goal and had enrolled only a little more than half of its target by the end of the cooperative agreement. This was mainly due to two key challenges. The first was difficulty obtaining referrals to the program. Clinicians who referred patients to the program indicated that it was challenging to determine whether a patient met the eligibility criteria. The second challenge was a lower participant acceptance rate than expected. An internal evaluation of the program found that one-third of all eligible patients refused to participate (Federman et al. 2018). Clinicians and staff said that issues related to the amount of time and the burden of other requirements needed to participate, patients' satisfaction with the traditional system of care, and privacy and confidentiality might have contributed to patients' refusals. Mount Sinai addressed these challenges by changing the eligibility criteria and recruitment processes. For example, it relaxed the inclusion criteria to allow MACT physicians more clinical judgment in determining which patients the MACT could safely treat at home regardless of whether they had one of the original eight conditions required by the program.[2] The MACT also hired and trained physician assistants who already worked in the Mount Sinai ED to moonlight for the HaH program to expand recruitment hours and leverage their established relationships with ED doctors.

[2] The conditions are congestive heart failure or heart failure, chronic obstructive pulmonary disease or asthma, dehydration, diabetes, pneumonia, cellulitis, urinary tract infection, and pulmonary embolism or deep vein thrombosis.

The HaH program also had challenges obtaining adequate services from outside vendors. Mount Sinai originally designed the HaH program to rely primarily on outsourced services to demonstrate that a variety of organizations could implement the program as long as they had established contracts with other organizations to provide any services that were not available from the primary organization. As such, Mount Sinai contracted with another organization to provide nursing to HaH patients because Mount Sinai did not have nursing staff to provide home infusion and other home-based skilled nursing services. However, many outside vendors were accustomed to working normal business hours and had difficulties following the around-the-clock schedule necessary for the HaH program to provide services in the home. In response, Mount Sinai reduced its reliance on external contractors by hiring nurses internally to provide HaH services. These changes resulted in only Mount Sinai staff comprising the core MACT in the final year of the cooperative agreement. With this consistent group of Mount Sinai staff, the program could provide services more consistently and achieved a higher level of communication between the nurses and the rest of the MACT staff.

Despite the difficulties with program recruitment and enrollment, the awardee implemented the program on time and delivered services as intended. Overall the awardee recruited, hired, and retained staff throughout the cooperative agreement; however, staff noted a desire for additional trainings. Patients participated actively in the HaH program, engaging with MACT staff for the duration of their care. Clinicians reported the program had a positive effect on the access to and quality of care they provided to participants, as well as participants' satisfaction and quality of life. Mount Sinai reported that, among respondents to the Hospital Consumer Assessment of Healthcare Providers and Systems survey, HaH participants assigned higher ratings to communication with physicians and with nurses than did those eligible nonparticipants.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Recruiting, enrolling, and engaging participants

MACT administrative assistants reviewed the hospital's electronic medical records to identify potentially eligible patients during an ED visit. After identifying a potentially eligible patient (or one referred into the program by a provider) in the ED, the MACT physician reviewed the patient's clinical information to assess whether the patient was appropriate for the program. If so, the physician discussed the HaH program as an option with the patient's care team. Patients were eligible for the program if they were covered by a participating health plan, including Medicare FFS, and if they met the MCG for hospital admission. MCG are a set of evidence-based guidelines that help to support decision making related to admitting patients to the hospital. The program excluded patients if they did not live in Manhattan, were clinically unstable, or needed cardiac monitoring or intensive care. Clinical judgment played a large role in determining a patient's suitability for participation in the HaH program. Clinical staff also assessed the patient's home environment during the initial interview in the ED and again when they accompanied the patient from the hospital to the home. Staff assessed the overall safety of the

home, the availability of family or other support members to assist with meals and telephone contact, and accessibility to an unshared bathroom.

The MACT team engaged most patients who enrolled in the HaH program for the duration of their care. Patients and their caregivers had to engage in the patients' care because many of the activities—such as helping the patient to the bathroom or preparing food, which nurses or support staff would have completed in an inpatient setting—were left to the patient and his or her support system to complete. Most MACT clinicians (87 percent) who provided direct patient care and responded to the clinician survey strongly agreed that the MACT team had successfully engaged patients with the program.

Characteristics of Medicare FFS participants in the HaH program

The awardee enrolled 295 participants, but only 184 beneficiaries (62 percent) of the awardee's list of participants were enrolled in Medicare FFS and met the claims-based eligibility criteria for the study from November 2014 through August 2017 (Appendix A, Table A.1). This subset of participants was predominantly an elderly group (Table 3). More than three-quarters of the recruited beneficiaries were ages 75 or older (78 percent) nearly three-quarters (72 percent) were female. Most participants were originally eligible for Medicare based on age (84 percent), whereas 15 percent were eligible because of a disability. In addition, 30 percent were dually eligible for Medicare and Medicaid. This indicates a high level of social need, considering that 18 percent of beneficiaries nationwide are dually eligible. Overall, participants in the HaH program were substantially less healthy and had a greater need for care than the general Medicare FFS population, as evidenced by the fact that the average hierarchical condition category (HCC) risk score for participants (2.8) was almost triple the average score for Medicare FFS beneficiaries nationwide (1.0).

Table 3. Baseline characteristics of Medicare FFS beneficiaries

Characteristic	Medicare FFS participants (N = 184)
Demographics, %	
Age group	
Younger than 65	7
65 to 74	15
75 to 84	27
85 and older	51
Female	72
Male	28
White	67
Black	18
Hispanic	10
American Indian, Alaska Native, Asian/Pacific Island American, or other	3

Table 3 (*continued*)

Characteristic	Medicare FFS participants (N = 184)
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	84
Disability insurance benefits	15
ESRD[a]	1
Medicare/Medicaid dual status, %	**30**
Dual eligible	30
HCC score[b]	
Mean	2.75
25th percentile	1.47
Median	2.29
75th percentile	3.63
Service use and expenditures during the year before enrollment	
Any hospitalizations, %	58
Any outpatient ED visits, %	78
Percentage with a 30-day readmission among all discharges, %	23
Total Medicare expenditures ($ PBPM)	3,167
Acute inpatient expenditures ($ PBPM)	1,521
Number of outpatient ED or OBS visits (per 1 ,000)	1,651
Number of hospital admissions (per 1,000)	1,188

Sources: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data as of August 31, 2017.

Notes: The evaluation defined the baseline year as the 365 days before each beneficiary's enrollment date. It defined the enrollment date as the start of the beneficiary's HaH care at home. The evaluation measured all beneficiary characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

[a] Includes participants with both a disability and ESRD.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HaH = Hospital at Home program; HCC = hierarchical condition category; OBS = observation stay; PBPM = per beneficiary per month.

Consistent with their high needs, participants had high rates of Medicare expenditures and service use in the year before enrollment (Table 3). The total average per beneficiary per month (PBPM) Medicare payment during the baseline year was $3,167, which was substantially higher than the 2014 national average of $792.[3] The average PBPM Medicare payment for inpatient care ($1,521) was the largest driver of the total cost of care, representing almost half of this cost

[3] The national data here and in the next paragraph are from the Centers for Medicare & Medicaid Services. "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

(48 percent). In addition, more than three-fourths of the participants (78 percent) had an ED visit, resulting in an annual rate of outpatient ED visits of 1,651 per 1,000 participants during the baseline year. More than half (58 percent) of participants had a hospital admission in the baseline year, leading to an annual rate of acute hospital admissions of 1,188 per 1,000 participants. Similarly, the 30-day unplanned readmission rate for participants (23 percent) was much higher than the national rate of 18 percent per discharge.

Challenges of estimating program impacts

It was not possible to conduct a rigorous impact evaluation of the HaH program and measure the impact of the intervention on service use and costs among Medicare FFS beneficiaries for two reasons. First, the HaH eligibility criteria for enrollment relied on clinical information that was not available in claims or other administrative data. For example, when the program identified or referred a potentially eligible patient, the MACT physician reviewed the patient's clinical information to assess whether the patient was appropriate for the program and discussed the MACT program as an option with the patient's care team. Such clinical assessments cannot be replicated using Medicare FFS enrollment and claims data. In addition, program eligibility required a safe and supportive home environment, information that could not be ascertained for a comparison group. Therefore, it was not possible to identify from claims data a comparison group that met the HaH eligibility criteria. Second, because only 184 of the 295 patients enrolled in HaH program could be identified in the Medicare enrollment database and met the claims-based eligibility criteria, it is highly unlikely that an impact evaluation could obtain statistically significant impact estimates even if true program effects were quite large (for example, 20 percent of expenditures). For these reasons, this report does not present impact estimates.

CONCLUSION

Mount Sinai implemented the HaH program on time and provided acute and post-acute care services to patients in the home. Most staff who provided direct care and responded to the nonclinician or clinician surveys perceived that the program had positive impacts on the delivery of care and patients' outcomes. Despite these successes, the HaH program had difficulties meeting its initial enrollment goal and had enrolled only a little more than half of its target by the end of the cooperative agreement. In addition, although originally designed to rely primarily on outsourced services, the HaH program reported challenges obtaining adequate services from outside vendors. In response, Mount Sinai reduced its reliance on external contractors by hiring nurses internally to provide HaH services. Because the evaluation could not replicate the eligibility criteria using Medicare claims data or clinical registry data, it was not possible to identify a comparison group that was similar to the intervention group at the time of enrollment into the program. That, and the small sample size, made it impossible to conduct a rigorous impact evaluation of this program.

PROGRAM SUSTAINABILITY

Since its award ended August 2017, Mount Sinai continued the MACT program by providing a slightly modified program to some patients but not others, based on their source of insurance. Medicare FFS beneficiaries stopped receiving MACT services due to a lack of federal funding. Mount Sinai continued to provide MACT acute and post-acute services to patients covered by two commercial payers with which the awardee had executed contracts. The awardee modified program services to align with these payers' requirements. The modifications included (1) limiting HaH services only to patients treated as inpatients or who received palliative care, and no longer including patients who preferred not to be admitted to the hospital; (2) expanding participant eligibility criteria to include eight zip codes in two New York City boroughs chosen because of their proximity to Mount Sinai hospitals and accessibility to public transportation; and (3) formalizing some MACT processes, such as developing a protocol for managing adverse events.

Icahn's proposed payment model

To fund the HaH program of MACT, Mount Sinai developed a bundled payment model that it could tailor based on the payer's preferences. The bundle of services included the core MACT services in the acute care phase that did not already have a payment mechanism. The program could expand the bundled payment model to include the post-acute care phase of the HaH program and could incorporate a shared risk component.

The awardee also developed a separate Medicare payment model that included a bundled payment with risk sharing. The payment model covered program services for three of the program's target populations: patients receiving acute care, palliative care, and observation unit services. The payment would cover MACT services in the acute and post-acute phases, including the MACT doctors' professional fees, nursing care, social work services, and community paramedicine visits. Services not covered in the bundle (such as inpatient consultations and post-acute radiology services) would be billed separately as FFS.

Finally, the HaH provider would be eligible for shared savings or responsible for repaying a percentage of the annual difference between the total spending in HaH episodes and the benchmark.

Despite successfully executing contracts for its bundled payment model with two commercial payers, Mount Sinai reported concerns that it could not sustain the program without a Medicare payment model. Medicare FFS beneficiaries make up a large portion of patients eligible for the program, and payments from commercial payers are unlikely to cover the ongoing, long-term costs of the program. Mount Sinai had proposed a Medicare payment model to the Physician-Focused Payment Model Technical Advisory Committee, which had recommended it to the Secretary of the U.S. Department of Health and Human Services for implementation. However, the Secretary indicated that the agency would not implement the model (Azar 2018).

REFERENCES

Azar, Alex M. "The Secretary of Health and Human Services' Response to Comments and Recommendations of the Physician-Focused Payment Model Technical Advisory Committee Transmitted Between October 2017 and May 2018." Washington, DC: U.S. Department of Health and Human Services, June 13, 2018. Available at https://aspe.hhs.gov/system/files/pdf/255906/Secretarial_Responses_June_13_2018.508.pdf. Accessed August 17, 2018.

Federman, Alex, Tacara Soones, Linda DeCherrie, Bruce Leff, and Albert L. Siu. "Association of a Bundled Hospital-at-Home and 30-Day Postacute Transitional Care Program with Clinical Outcomes and Patient Experiences," *JAMA Internal Medicine,* vol. 178, no. 8, June 25, 2018, pp. 1033–1040.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants excluded from analytic sample	Number of participants remaining in analytic sample
Total program participants		**295**
Not enrolled in Medicare or could not be identified in Medicare enrollment files	50	245
Lack of Part A and B enrollment on date of HaH program enrollment	5	240
Enrolled in Medicare Advantage	52	188
Medicare was not primary payer	1	187
Insufficient FFS enrollment period at baseline	3	184
Final analytic sample		**184**

Sources: Mathematica's analysis of information from the awardee's program encounter database from November 1, 2014, through August 31, 2017, and Medicare claims and enrollment data from November 2013 through February 2018, as of October 2019.

FFS = fee-for-service; HaH = Hospital at Home program.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Johns Hopkins University

September 2020

Margaret O'Brien-Strain, Elizabeth Combs, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

JOHNS HOPKINS UNIVERSITY

Johns Hopkins University received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to support the Maximizing Independence at Home (MIND) program. The purpose of the MIND program was to identify and address the unmet needs of individuals diagnosed with Alzheimer's disease or another dementia-related neurodegenerative disease and the unmet needs of their caregivers. The MIND program sought to delay or prevent participants from moving out of their homes and into supported living facilities or nursing homes and to reduce hospital and emergency department (ED) admissions. The target population consisted of older adults living in and around Baltimore, Maryland, who had been diagnosed with or suspected of Alzheimer's disease or another dementia-related neurodegenerative disease. The intervention originally targeted Medicare–Medicaid dual eligible individuals but expanded to include Medicare-only beneficiaries. The intervention launched in March 2015. The intended period of service for the MIND intervention was 18 months. The awardee ended enrollment in September 2016 to maximize the share of participants who received the full intervention. Johns Hopkins University received a three-month no-cost extension, and its program ended November 30, 2017. Table 1 summarizes the program's key characteristics.

The awardee hypothesized that individualized in-home care management services would enable participants to remain longer at home. Memory care coordinators (MCCs) worked with an interdisciplinary team—composed of a geriatric psychiatrist, an occupational therapist, and registered nurses (RNs)—to address participants' needs through regular interactions and home visits. The MCCs (1) connected participants to meaningful activities, (2) educated caregivers in how to mitigate participants' risky behaviors, (3) referred participants to services to improve the safety of the home environment, and (4) worked closely with the clinical team to coordinate care.

Important issues for understanding the evaluation

- The MIND program enrolled Medicare and Medicare–Medicaid dual eligible individuals with dementia. The program enrolled 342 participants and their caregivers.
- Due to the small study sample and an inability to replicate the eligibility criteria in claims, it was not possible to conduct a rigorous impact evaluation of this program.
- This descriptive analysis relied on the awardee's self-measurement and monitoring data for 342 participants (and service use and expenditure data on a subset of 249 Medicare fee-for-service [FFS] beneficiaries) and should therefore not be interpreted as an evaluation of program impacts.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Johns Hopkins University implemented the MIND at Home program to identify and address the unmet needs of adults who had been diagnosed with Alzheimer's disease or another dementia-related neurodegenerative disease and their caregivers to improve health outcomes, reduce health care costs, and prevent or delay institutionalization.
Major innovation	The MIND program was innovative in using individualized care management and caregiver support delivered in the participant's home by MCCs with the support of an interdisciplinary team.
Program components	• Care management to address unmet needs of individuals with dementia • Patient and family engagement to support family caregivers to delay institutionalization • Training and health technology to support home health agency staff as MCCs
Target population	The program sought to engage Medicare beneficiaries in and around Baltimore, Maryland, who had been diagnosed with or suspected of Alzheimer's disease or a related form of neurodegenerative dementia.
Total enrollment	Johns Hopkins University enrolled 342 participants and their caregivers in the MIND program (114 percent of its revised enrollment projection) by the end of the three-year cooperative agreement.
Theory of change or theory of action	By addressing participants' unmet social and medical needs, supporting caregivers, and improving home safety, the program aimed to help participants stay in their homes longer and reduce costs of nursing home care. Additional support for caregivers could reduce ED visits and hospital admissions.
Award amount	$6,387,736
Effective launch date	The program began operating on March 2, 2015, six months after the award date.
Program settings	MCCs delivered program services in participants' homes.
Market area	Baltimore, Maryland and the surrounding area
Target outcomes	• Reductions in unmet needs and improved quality of life for patients • Cost savings due to delay in transition to nursing home • Cost savings due to fewer ED visits and shorter hospital stays
Payment model	By the end of the award, Johns Hopkins University had developed a payment model with PBPM payment shared between the MCCs and the clinical team. However, the awardee had been unable to negotiate agreements with payers that integrated care between Medicare- and Medicaid-funded nursing home care.
Sustainability plans	The MIND at Home model was developed outside an existing health system, so it did not continue after the award period. However, Johns Hopkins University has developed web-based training modules and was exploring strategies for telementoring approaches.

ED = emergency department; MCC = memory care coordinator; MIND = Maximizing Independence at Home; PBPM = per beneficiary per month.

It was not possible to conduct a rigorous impact evaluation of the MIND program because of the way in which the awardee identified participants and the low number of enrollees. As a result, this report describes only the demographic and health characteristics of Medicare participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis. Appendix A, Table A.1 describes the identification of the sample used for the descriptive analysis.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	A rigorous impact evaluation of this program was not possible because the awardee used eligibility criteria that claims data do not contain, precluding the identification of a credible comparison group, and because the awardee did not enroll enough participants with sufficient exposure to intervention services to detect impacts of reasonable magnitude. The evaluation consists of a descriptive analysis of Medicare FFS beneficiaries.
Intervention group for descriptive analysis	The intervention group for the descriptive analysis comprised the 249 participants (among the total enrollment of 342) who could be identified as being enrolled in Medicare FFS Parts A and B with Medicare as a primary payer and who had at least 90 days of Medicare claims history before enrollment into the program.
Limitations	Because this study could not identify a comparison group, no impact estimates were constructed.

FFS = fee-for-service.

PROGRAM DESIGN AND ADAPTATION

Johns Hopkins University designed the MIND program around the MCCs, supported by an interdisciplinary clinical care team.[1] The MCCs helped participants reduce unmet social, medical, and safety needs by (1) connecting them to meaningful activities and services, (2) educating caregivers on how to mitigate participants' risky behaviors, and (3) coordinating social and medical services. Coordinating services included reminding participants to schedule necessary appointments, arranging transportation, and monitoring adherence to medication regimens. By keeping participants safer and happier in their homes, the MIND program attempted to delay or prevent participants from moving into institutions and to reduce ED visits and hospitalizations.

In the last program year, Johns Hopkins University expanded the array of services offered through the MIND program by adding tele-health consultations, occupational therapy, and a monthly newsletter. Tele-health connected participants and their caregivers with a geriatric psychiatrist to discuss clinical needs and next steps in care. Johns Hopkins University also piloted occupational therapy services through its partnership with the Tailored Activity Program (TAP), a home-based occupational therapy intervention for individuals with dementia and their caregivers. Although the services provided through TAP were beneficial to some MIND participants, the awardee felt that the services were not a good fit for the more intensive needs of MIND participants. Finally, the awardee introduced a monthly newsletter to help area organizations working with older adults to share information about related resources and events.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of MIND program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Johns Hopkins University was effective in implementing the MIND program and faced few barriers to program implementation, other than enrollment challenges. The program contracted with MCCs through the partner home health agencies and the MIND project team—a geriatric psychiatrist, RN, and an occupational therapist—supported them. The MCCs received formal and informal training through the MIND program. New MCCs also shadowed more experienced MCCs to gain tips on client interactions and identify available services and supports that could address clients' needs. Finally, in weekly meetings with the project team, MCCs worked through a participant's care plan in a group setting and learned more about caring for people with dementia. Among surveyed staff, the majority (80 percent) reported that the training helped them learn new skills and improve job performance. The MCCs were valued for being resourceful, persistent, and not becoming overwhelmed by the challenges their participants faced, but instead prioritizing and calmly tackling issues one at a time.

Beyond the success of the MCCs, two other strengths of the implementation were the partnership with home health agencies and the use of the Dementia Care Management System (DCMS), a health technology system designed for the MIND program. If adopted more broadly, a MIND model would likely operate out of home health agencies. However, these agencies did not play a large role in developing and implementing the MIND program, beyond contracting the MCCs. Nonetheless, home health agency leaders reported that the MIND program was a natural fit for their organizations. The DCMS was an important asset both in tracking participants' needs, progress, and outcomes, and as a repository of resources for the MCCs. MCCs and program staff populated the DCMS with organizations, service providers, and tips for treating dementia symptoms and behaviors.

Low enrollment was the primary challenge for the MIND program. Two major issues were limited success recruiting Medicare–Medicaid dual eligible beneficiaries and underdiagnosis of prior dementia, two of the original eligibility criteria. Given its main goal of delaying transitions to nursing home care, the greatest opportunity for public insurance savings arose from reducing or delaying Medicaid-funded institutional care. However, Johns Hopkins University had difficulty recruiting enough dual eligible participants to generate meaningful savings. Program leaders believed this arose in part because of

Implications of program implementation for achieving program goals

- MCC's received formal and informal training and met weekly with the project team to strengthen their ability to provide care coordination.
- A dementia care case management system and partnerships with home health agencies were critical components of the model.
- Low enrollment was the primary challenge and required the awardee to expand eligibility criteria.

oversaturation of research in Baltimore and community mistrust of research and the consent process. Johns Hopkins University cited the severe social isolation that is common among dual eligible beneficiaries with dementia. The program also discovered that dementia tended to be underdiagnosed among the patient population living in Baltimore and the surrounding counties. As a result, many potential participants did not have a prior dementia diagnosis.

Faced with these challenges, the MIND program cut its initial enrollment target in half, from 600 to 300, and changed its eligibility criteria. First, it expanded the program to include Medicare-only beneficiaries. Second, it expanded its catchment area beyond Baltimore to surrounding counties. Third, it allowed individuals with *suspected* dementia to enroll in the program. After an initial phone screen, a MIND geriatric psychiatrist assessed these individuals and provided a formal diagnosis when applicable.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

The descriptive analysis in this report presents findings on enrolling and engaging participants and implementing the program based on Johns Hopkins University's self-measurement and monitoring reports. Medicare claims for 185 FFS beneficiaries who were enrolled in the MIND program for at least 18 months supplement that information, as do interviews and surveys with MIND program staff.

Recruiting and enrolling participants

The awardee enrolled 342 participants by September 2016, when enrollment stopped to provide sufficient time for the intervention. These participants represented 114 percent of the revised target, but 57 percent of the projected enrollment. Slow enrollment in the first program year prompted Johns Hopkins University to shorten the intervention period from 26 to 18 months. Two-thirds of participants (230) enrolled in the last six months of the enrollment window, including 93 who received less than the full 18 months of service before the end of the program in November 2017.

The MIND program was successful in engaging and retaining participants and in reducing their unmet needs. Based on the awardee's self-reported process metrics, MCCs made an average of 4.3 contacts per participant per month at peak participation (the 18 months from March 2016 to August 2017), including 3.3 telephone contacts, 0.7 in-person meetings, and 0.4 written contacts per month. Over an 18-month intervention, this totaled 78 contacts, with 59 phone calls and 13 in-person meetings per participant. The MCCs reported that the mode and frequency of contacts reflected the individual patient's needs, hospitalizations, and living situation. The MCC interviewees reported that some participants needed more attention than others. Participants' needs also changed over time as they faced and resolved various challenges or crises. For example, a participant going through a medical crisis needed the MCC's frequent support in communicating with clinicians or addressing hospital discharge needs.

As measured by the Johns Hopkins Dementia Care Needs Assessment (JHDCNA), unmet needs for people with dementia declined during the intervention. Based on data reported by the awardee, the number of unmet needs fell 38 percent on average between the first assessment at enrollment and the final assessment in Month 18. There were decreases across all domains of the assessment, with the largest changes in meaningful activities, home and personal safety, and neuropsychiatric assessment. Needs for care financing saw the smallest decrease but were still 21 percent below the baseline assessment by Month 18. Findings from a program-administered care coordination satisfaction survey were consistent with the JHDCNA results. About 95 percent of caregivers reported that they would recommend MIND, and they reported that they felt more confident, in control, and educated about dementia as a result of the program. Among other themes, Johns Hopkins University reported that caregivers appreciated access to knowledgeable staff (including clinicians who visited their home), the personalized care and recommendations, and the support from "the care coordinators to help them feel they are not alone."

By the end of the program in November 2017, 58 of the 342 participants had been lost to follow-up, withdrawn, or did not complete data collection. Of the remaining 284 participants, 41 (14 percent) had permanently transitioned to a higher level of care (such as assisted living, memory care, or nursing home care). Another 48 participants (17 percent) had died. That left 195 (69 percent) of the 284 participants with a known status remaining at home.

Characteristics of program participants

Johns Hopkins University's self-reported enrollment statistics show that the MIND program reached a relatively high-need population (Table 3). The average age at enrollment was 81, and most had moderate or severe dementia, as determined by the JHDCNA. The program required participation by a caregiver, and two-thirds of enrollees lived with the caregiver. The program assessed enrollees' dementia stage using the Mini-Mental State Exam, with 46 percent of enrollees scored as moderate dementia and 16 percent as severe dementia. At enrollment, the initial assessment found an average of 12 unmet needs, out of a possible 43.

Among the 342 total enrollees, Medicare FFS data were available for 249. The analysis excluded 93 participants because they could not be linked to the Medicare enrollment database (28), were not enrolled in both Parts A and B (3) or were enrolled in Medicare Advantage (59), did not have Medicare as a primary payer (1), and did not have at least 90 days of Medicare claims history before enrollment into the program. (Appendix A, Table A.1 describes the identification of the sample used for the descriptive analysis.) The demographics of this Medicare FFS subset essentially matched the participant population overall. Three-quarters of the participants were women. Slightly more than 60 percent of enrollees identified as Black or African American, 31 percent as White, and 8 percent as other races or ethnicities. Of the 249 participants in the analysis data, 70 percent were Medicare–Medicaid dual eligible beneficiaries. The average hierarchical condition category (HCC) risk score for MIND Medicare FFS beneficiaries were nearly twice the average Medicare FFS HCC score nationally indicating that they are expected to have much higher costs than the average Medicare FFS beneficiary.

The high level of health care service use and Medicare spending during the year before enrollment also reflect the substantial health care needs of MIND participants. The total average per beneficiary per month (PBPM) Medicare expenditure during the baseline year was $1,856, more than twice the 2014 national average for Medicare FFS beneficiaries of $792. This is consistent with the average HCC score of about 2 and the rate of hospitalizations per 1,000, which is also over twice the national average.

Table 3. Baseline characteristics of MIND participants

Characteristic	All participants (N = 342)	Medicare FFS participants (N = 249)
Demographics		
Age at enrollment, years	81	81
Female, %	75	74
Race, %		
White	31	31
Black	61	60
All other[a]	8	9
Awardee-reported characteristics at enrollment		
Living with caregiver, %	68	
Dementia stage, %		
Mild	36	
Moderate	46	
Severe	16	
Total unmet needs (out of 43), mean	12.3	
Original reason for Medicare eligibility, %		
Old age and survivor's insurance		78
Disability insurance benefits[b]		22
Medicare–Medicaid dual status, %		70
HCC score[c]		
Mean		1.99
25th percentile		1.05
Median		1.58
75th percentile		2.53
Service use and expenditures during the year before enrollment		
Any hospitalizations, %		37
Any outpatient ED visits, %		50
Number of hospital admissions (per 1,000)		688
Number of outpatient ED visits (per 1,000)		1,011
Total Medicare expenditures ($ PBPM)		1,856

Sources: Data for the all-participants population came from awardee's performance dashboard through November 30, 2017. Data for the Medicare FFS population came from Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data as of July 31, 2017.

Table 3 (*continued*)

Note: The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

[a] Other racial or ethnic groups are combined for comparability between Medicare data and awardee reports.

[b] Includes participants with both a disability and an ESRD.

[c] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; MIND = Maximizing Independence at Home program; PBPM = per beneficiary per month.

Challenges of estimating program impacts

A rigorous evaluation of the MIND program's impacts on time to nursing home care and Medicare–Medicaid expenditures and service use was not possible for four reasons. First, there were too few beneficiaries in the intervention group to detect effects for primary outcomes. Second, the major effect on expenditures was expected to arise from lower Medicaid costs from slowing down transitions to nursing home care for Medicare–Medicaid dual eligible beneficiaries. The awardee was unable to recruit enough dual eligible beneficiaries to meet even its reduced target enrollment and, thus, the program extended eligibility to Medicare-only beneficiaries, whose nursing home care would not be paid by Medicaid in the observation period. Third, it was not possible to construct a comparison group from claims data because the MIND program also extended enrollment to individuals with no prior dementia diagnosis, and screening for dementia was part of the enrollment process. Such screening identified individuals with dementia at an earlier stage than would occur in the absence of the program. For this reason, individuals with dementia diagnoses in claims are likely to be at later stages of dementia than MIND participants. Finally, MIND participants had to have an involved caregiver, usually living in the home of the participant. The proximity and involvement of a caregiver is likely to affect outcomes, especially transition to nursing home care. This criterion also could not be replicated in claims. The results presented here therefore cannot be interpreted as an evaluation of program impacts.

CONCLUSION

Johns Hopkins University succeeded in implementing its MIND program to provide memory care coordination for individuals with dementia and their caregivers. By the end of the three-year cooperative agreement, the awardee had enrolled 114 percent of its revised enrollment target (57 percent of its original projection); trained MCC's to support individuals through in-person, telephone, email, text, and mail messages; and delivered services in a manner consistent with the design proposed in its HCIA R2 application. The awardee successfully broadened the eligibility criteria to substantially increase enrollment in the last six months of its enrollment window. The awardee's self-monitoring statistics indicate high levels of participant contacts, reductions in unmet needs, and high rates of participants' satisfaction with the MIND program. At the end of the program, 69 percent of individuals with known participation status continued to reside at home. However, a rigorous impact evaluation of this awardee was not possible due to individuals

qualifying for the program for reasons not observable in Medicaid claims data, and the small size of the sample.

PROGRAM SUSTAINABILITY

Johns Hopkins University did not continue the MIND program after its award ended in November 2017. Because the services were not provided within an existing health care system, no avenue for sustainability existed at the end of the award period. Nonetheless, Johns Hopkins University continued to develop specific components of the model after the award ended. In particular, it developed the MIND training into web-based training modules that were complete and ready for use. As of November 2017, the awardee was also working with a Johns Hopkins University work group on telemedicine, exploring strategies to replicate MIND at Home using a telementoring approach similar to that employed by the University of New Mexico's Project ECHO (https://echo.unm.edu/data).

Johns Hopkins University's proposed payment model

Johns Hopkins University proposed a payment model that would provide a $250 PBPM payment for memory care coordination, shared between MCCs and a clinical team. The fee could potentially be risk-adjusted to reflect the intensity of care needs.

Because savings arise from delayed nursing home care paid by Medicaid, example target organizations for the program payment would be Medicaid ACOs, Integrated Care Organizations, PACE programs, and similar systems or providers that bridge Medicaid and Medicare. By the end of the award, the awardee had been unsuccessful in negotiating agreements with payers.

In addition, the awardee developed and proposed a payment model that would have provided a PBPM care coordination payment for MIND services, split between the MCCs (located within a home health agency) and the clinical team (located within Johns Hopkins University), and payable as long as a participant remained living at home. The awardee's proposed PBPM fee was $250, although Johns Hopkins University intended to conduct additional analyses of ways to have a risk-adjusted fee based on the intensity of memory care services needed.

One challenge with this payment model was that Medicaid was the payer likely to benefit from delayed entry to nursing homes, and Medicare would have been the likely payer for the MIND program. For this reason, the awardee explored partnerships with providers in financial or care settings that bridged these payment systems. Examples include the Financial Alignment Initiative for Medicare–Medicaid Enrollees, Medicaid Accountable Care Organizations, Integrated Care Organizations, PACE programs, continuing care retirement communities, health systems, and home care agencies. At the end of the award period, the awardee was developing a business case for the model but had not negotiated an agreement with payers.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total Medicare beneficiaries in awardee's finder file as of August 31, 2017		**342**
Could not be linked to Medicare enrollment database	28	314
Did not meet study's standard claims-based inclusion criteria		
Not enrolled in both Part A and B	3	311
Enrolled in Medicare Advantage	59	252
Medicare not primary payer	1	251
Fewer than 90 days of claims history before enrollment	2	249
Final Medicare FFS beneficiaries in descriptive analysis		**249**

Source: Mathematica's analysis of information from awardee's finder file and Medicare claims as of August 31, 2017.
FFS = fee-for-service.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Montefiore Medical Center

September 2020

Jelena Zurovac, Kate Stewart, Stacy Dale, Danielle Chelminsky, and Jennifer Lyons

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

MONTEFIORE MEDICAL CENTER

Montefiore Medical Center (Montefiore), a large tertiary care center in the Bronx, New York, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the Behavioral Health Integration Program (BHIP) in a subset of its 22 primary care sites. The BHIP is based on a measurement-based collaborative care model, in which primary care physicians (PCPs) and behavioral health staff work together in the primary care setting to provide behavioral health care services and referrals. The awardee aimed to screen all adults seen at participating sites for depression, anxiety, and alcohol use disorders, and all children and adolescents for attention deficit hyperactivity disorder (ADHD). After screening positive and being evaluated by the PCP, participants with one of the targeted conditions received behavioral health care services. Behavioral health staff used data from initial and follow-up assessments to monitor participants' progress and adjust their approach to treatment according to participants' changing needs. Three primary care sites started enrolling participants in February 2015, and three more in August 2015. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program aimed to improve health and reduce emergency department (ED) service use, inpatient admissions, and costs by integrating behavioral health and primary care services.
- This impact analysis relied on 2,069 Medicaid beneficiaries who, after screening positive for depression, received BHIP services and 1,432 comparison beneficiaries at nonparticipating sites who screened positive for depression and had similar demographic and health characteristics.

  Because nonparticipating sites also provided behavioral health services, the evaluation could assess only the additional effect of the more intensive services (team-based care and ongoing patient monitoring) provided at participating sites.

Awardee leaders hypothesized that the program would increase participants' satisfaction with care, improve their behavioral and somatic (physical) health, and reduce the costs of care. Specifically, the awardee posited that alleviating an individual's depression and anxiety symptoms often leads to better self-care and self-management, better adherence to lifestyle changes and medications, and better physical health. The program aimed to improve health and reduce emergency department (ED) use, inpatient admissions, and the costs of care.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Montefiore Medical Center provided behavioral health screening and treatment services to patients in 6 of its 22 primary care sites in the Bronx to improve health outcomes and reduce the cost of care.

Table 1 (*continued*)

Program characteristics	Description
Major innovation	Montefiore implemented the collaborative care model in a high-need population, using technology platforms to improve patient engagement and service delivery.
Program components	• Integrated behavioral health and primary care services • Health IT, including patient registry to collect data on participants and monitor their progress through measurement-based care and telemedicine tools to administer follow-up measures and to engage and communicate with participants
Target population	Patients who received services from participating primary care sites and screened positive for depression, anxiety, risk of alcohol use disorders, or (for children and adolescents) ADHD
Participating providers	6 primary care sites participated in the program: 3 sites started in February 2015 and three in August 2015.
Total enrollment	A total of 6,559 patients enrolled in the program from February 2015 through August 2018, 143 percent of the awardee's original enrollment goal. The total number of participants included those who screened positive for depression, anxiety, alcohol use disorder, or (for children and adolescents) ADHD.
Level of engagement	By the end of the first year of program implementation, participants had four visits on average with a behavioral health staff member. By the end of the third year of the program, participants had about seven visits on average during their time in treatment, which averaged 13 weeks.
Theory of change or theory of action	PCPs who work with on-site behavioral health staff to measure and respond to participants' progress together addressed participants' behavioral health needs that might otherwise go untreated. The awardee expected that improved access to behavioral health care (that is, integrated with primary care services) would lead to increased satisfaction with care, better physical and behavioral health outcomes, fewer hospitalizations, and lower costs.
Award amount	$5,583,090
Effective launch date	• Cooperative agreement period started in September 2014 • The program began operating in February 2015, when three primary care sites began to enroll participants. Three more sites began enrolling patients in August 2015.
Program settings	Six primary care practices within the Montefiore Medical Center
Market area	Urban (the Bronx, NY)
Target outcomes	• Increase in patients' satisfaction • Improvement in participants' behavioral health and chronic disease outcomes • Net savings in the cost of care for participants through fewer hospitalizations and ED visits
Payment model	Value-based payments; monthly bundled payment for care management and coordination services; shared savings through the Next Generation ACO
Sustainability plans	From September 2017 to the end of the cooperative agreement in August 2018, the awardee worked to expand the program beyond the 6 original health centers to its remaining 22 health centers. At the end of its cooperative agreement in August 2018, the awardee planned to sustain the BHIP through its payment model and state DSRIP payments.

ACO = accountable care organization; ADHD = attention deficit hyperactivity disorder; DSRIP = Delivery System Reform Incentive Payment; ED = emergency department; IT = information technology; PCP = primary care provider; PHQ-9 = Patient Health Questionnaire 9-Item.

This impact analysis was limited to 2,069 Medicaid beneficiaries who—after screening positive for depression and meeting the other study inclusion criteria —agreed to participate in the BHIP from May 2015 to August 2017. The comparison group included 1,432 Medicaid beneficiaries served by nonparticipating Montefiore sites who had similar demographic and health characteristics, who also screened positive for depression over the same period, and who were referred for treatment. Although the nonparticipating sites provided collocated behavioral health services, the services they provided did not involve team-based, collaborative care in which behavioral health teams monitored participants' progress and adjusted treatment based on reassessments maintained in a patient registry. In addition, at participating sites (but not at comparison sites), regular psychiatric case reviews were shared with the PCPs (Blackmore et al. 2018). As a result, the evaluation could assess only the additional effect of the more intensive services in participating sites versus the less intensive services in nonparticipating sites. Table 2 summarizes the key features of the impact evaluation. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study participants after versus before enrollment relative to the change in outcomes over the same period among a comparison group.
Intervention group for evaluation	Among 6,559 patients enrolled between February 1, 2015 and August 31, 2017, 2,450 were not Medicaid beneficiaries. The impact analysis relied on 2,069 Medicaid beneficiaries who—after screening positive for depression—agreed to participate in the BHIP from May 2015 through August 2017. The analytic sample excluded 1,574 treatment group beneficiaries who screened positive but did not meet the study's eligibility and claims-based inclusion criteria, such as not being enrolled in Medicaid for at least 90 days during the baseline year. The study also excluded 209 children, defined as those younger than 19 at the time of screening, 35 beneficiaries with missing PHQ-9 scores, 206 beneficiaries who enrolled during the program's first three months of operations when no comparison group members enrolled, and 16 treatment group beneficiaries with a PHQ-9 score of zero on their enrollment date.
Comparison group	The study also included 1,432 comparison beneficiaries diagnosed with depression who had similar characteristics. Comparison group beneficiaries also received collocated behavioral health services, though not as intensive and integrated as those in the treatment group.
Limitations	There is some chance for differential selection in that the treatment group beneficiaries screened positive and agreed to participate in the BHIP, whereas the comparison group beneficiaries were those who screened positive and were referred for services to the behavioral health care team. This likely resulted in more favorable estimated impacts than if comparison beneficiaries were as likely as participants to receive behavioral health care services

BHIP = Behavioral Health Integration Program.

PROGRAM DESIGN AND ADAPTATION

The BHIP centered on a measurement-based collaborative care model, in which behavioral health care was integrated into a primary care setting.[1] A key aspect of the program was maintenance of a patient registry that enabled clinic staff to monitor participants' progress over time and adjust treatment according to changing needs. BHIP staff also used several other health information technology (health IT) tools to screen patients, provide care, and monitor patients.[2]

Providing behavioral health services

The awardee aimed to administer screening for depression, anxiety, and alcohol use disorders to all adult patients seen at participating sites. The awardee then offered patients who screened positive for one or more conditions different combinations of behavioral health services, depending on their needs. Offered services included short-term psychotherapy with the licensed clinical social worker, psychiatric medication management by the PCP with support from the psychiatrist, and telephone outreach with behavioral activation and symptom monitoring from the behavioral health patient educator. The behavioral health team provided PCPs with regular psychiatric case reviews. Each site had a psychiatrist who consulted with (1) the behavioral health team to help identify participants' needs for behavioral health services and (2) the PCPs to support managing participants' psychiatric medications.

Health information technology

The behavioral health team conducted follow-up screening to monitor participants' progress and adjust treatment to changing needs. The behavioral health team used a patient registry to collect and track participants' screening scores, between-visit follow-up communications, and participants' care plans and goals. The patient educator or social worker referred participants whose screening scores did not improve during the program to the consulting psychiatrist for further assessment and recommendations for treatment.

In the second year of the cooperative agreement, the awardee incorporated telemedicine tools into the BHIP to boost participants' engagement with the program. Participants could subscribe to an interactive voice response technology that enabled them to complete follow-up monitoring measures via their phones, as well as to receive appointment reminders and health education messages. BHIP staff also rolled out a smartphone application that provided participants with follow-up behavioral health assessments; educational materials and videos; and reminders about treatment goals, appointments, and medications. The smartphone application also had a secure chat feature that enabled participants to communicate with patient educators. In addition, BHIP

[1] University of Washington, Advancing Integrated Mental Health Solutions. "Collaborative Care." Available at https://aims.uw.edu/collaborative-care.

[2] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

social workers could use a separate, Montefiore-wide, text messaging platform to receive and respond to messages from participants.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Despite challenges, Montefiore exceeded enrollment targets for adults. The awardee rolled out a new electronic medical record system shortly after the BHIP started, which caused short-term reductions in productivity and short-term lower-than-expected enrollment into the BHIP. The awardee also experienced ongoing challenges with staff turnover during the cooperative agreement period. Despite these challenges, the awardee exceeded its screening and enrollment targets for adults. For example, the depression screening rate among adults improved from 55 percent in fall 2015 (about six months after the BHIP started enrolling participants) to nearly 90 percent by summer 2017, when the last patients included in the analysis enrolled. The awardee experienced challenges screening children; for that reason, very few children were enrolled in the BHIP.

The awardee engaged a large proportion of adult participants, exceeded patient follow-up targets, and provided psychiatric consultations to a large proportion of participants who had not improved within the first three months of participation.

Social workers and case managers noted that interactive voice response and smartphone application enabled them to do their jobs more efficiently by providing automated means for collecting follow-up screening data and communicating with participants. These tools also improved their success rates in contacting participants.

Implications of program implementation for detecting impacts

- The awardee exceeded its enrollment, patient engagement, and follow-up targets for adult participants.
- Despite exceeding enrollment targets, the sample size is not large enough to detect small impacts.
- Because BHIP enrolled very few children, they are excluded from the analysis.
- Health IT enabled social workers and case managers to do their jobs more efficiently.
- The awardee implemented the BHIP with relatively few setbacks or challenges, which increased the probability of detecting impacts.

In the first two program years, awardee leaders and behavioral health staff reported that, although most PCPs expressed support for the program, PCPs were concerned about their availability for program activities. By the third year, PCPs were connecting more seamlessly with the program, in part due to improved workflow. Although PCPs continued to report competing priorities, they also said that they collaborated well with the behavioral health team and found value in BHIP services.

ESTIMATING PROGRAM IMPACTS

Study sample

The impact evaluation is based on 2,069 Medicaid beneficiaries who screened positive for depression and agreed to participate in the BHIP from May 2015 to August 2017, and 1,432 comparison beneficiaries who also screened positive for depression over the same period and were referred to behavioral health care services but were served by nonparticipating Montefiore clinics in the same region. We started with a sample of 6,559 participants. First, the study excluded 2,450 participants who were not Medicaid beneficiaries. The analytic sample also excluded 1,574 treatment group beneficiaries who did not meet standard eligibility and claims-based inclusion criteria for this study, such as not being enrolled in Medicaid for at least 90 days at baseline. The study also excluded 209 children (defined as those younger than 19 at the time of screening), 206 beneficiaries who enrolled during the program's first three months of operations when no comparison group members were enrolled, 35 beneficiaries with missing Patient Health Questionnaire 9-Item (PHQ-9) scores; and 16 beneficiaries with a PHQ-9 score of zero on their enrollment date. The study team applied the same exclusion criteria to the comparison group.

Because the awardee used health IT tools to screen patients, provide care, and monitor patients, the analysis could not separate the impact of health IT on outcomes from the effect of providing behavioral health services. For that reason, the study evaluates the effect of the intervention as a whole.

Due to data limitations, it was not possible to evaluate the effect of the intervention on BHIP participants who screened positive *only* for anxiety or alcohol use disorders. However, among BHIP participants who screened positive for depression, 82 percent also had moderate or severe anxiety symptoms (that is, Generalized Anxiety Disorder 7-item [GAD-7] score of 10 or higher).

Participant enrollment and recruitment

The awardee aimed to screen all adults who visited the BHIP's primary care sites during the award period for depression, anxiety, and alcohol use disorders. The PCP connected patients who screened positive to a member of the site's behavioral health team, composed of behavioral health patient educators and licensed clinical social workers. The process for connecting participants to BHIP staff varied by site and by clinician. It included in-person handoffs, phone referrals, or communication through the emergency medical record. The patient educator or social worker enrolled into the BHIP patients who agreed to participate by entering them into the program's patient registry.

Treatment group beneficiaries screened positive and agreed to participate in the BHIP, whereas the comparison group beneficiaries screened positive and were only referred to the behavioral health care team for services. This means that comparison beneficiaries might have been less likely to receive behavioral health services. If so, estimated impacts are more favorable than if both groups had the same likelihood of receiving services.

Participant engagement

The awardee engaged a large proportion of adult participants, exceeded patient follow-up targets, and provided psychiatric consultations to a large proportion of participants who had not improved within the first three months of participation. In the fall of 2015, the awardee reported engaging at least three-quarters of all patients who screened positive for one of the targeted conditions, which increased to 89 percent two years later. The BHIP aimed to provide at least three follow-up visits to participants within the first 10 weeks. Participants had nearly four follow-up visits on average during the first year of program implementation, and by the end of the third year, seven follow-up visits during their time in treatment, which averaged 13 weeks. Further, although only about one-third of participants who had not improved within three months received a psychiatric consultation in fall 2015; this increased to three-quarters a year later and to nearly all participants (94 percent) two years later.

Characteristics of treatment and comparison group beneficiaries

Comparing treatment and comparison group members at baseline confirmed that the two groups were well balanced in terms of their characteristics and health care service use (Table 3). Health care service use was similar for treatment and comparison beneficiaries during the entire baseline year and during the three months before enrollment. The average age of treatment and comparison group members during the baseline year was 40 years. Women comprised about 80 percent of participants and comparison group members, consistent with the fact that women are more likely than men to seek help for behavioral health conditions in primary care settings. Nearly all participants and comparisons lived in the Bronx or the five boroughs of New York City; very few lived in neighboring counties.

The PHQ-9 depression scores indicated an unmet need among participants for behavioral health services. Slightly more than half of participants had moderately severe or severe depression and slightly more than one-fifth had severe depression. Only 2 percent of participants in the analytic sample had minimal or no depression (PHQ-9 scores from 1 to 4). These patients were also likely to have enrolled in the program due to their anxiety. Nearly all participants with minimal or no depression had symptoms of moderate or severe anxiety, as evidenced by their GAD-7 scores of 10 and higher (data not shown).

The average Chronic Illness and Disability Payment System score was 1.9 for both participants and comparison group members, meaning that their expenditures were predicted to be almost twice as high as the average expenditures for adult Medicaid beneficiaries. Roughly 60 percent of both treatment and comparison group patients had a psychiatric condition at baseline. Also, both participants and comparison group members had several comorbidities that might also have caused them to use health care services, including emergency and hospital services. The next most common condition was cardiovascular disease, prevalent in about one-third of all patients, followed by pulmonary and skeletal conditions, prevalent in one-fifth of all patients. Appendix B provides the full balance results measured during the 12 months before enrollment.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Characteristic	Treatment mean (N = 2,069)	Comparison mean (N = 1,432)
Demographics		
Age at enrollment, years	40	40
Age group, %		
19 to 28	23	23
29 to 36	20	21
37 to 45	20	19
46 to 53	20	19
54 to 84	18	18
Male, %	21	20
Area of residence, %		
New York City (five boroughs) and Long Island	93	95
Outside New York City and Long Island	7	5
Medicaid enrollment, %		
Enrolled in comprehensive managed care	89	89
Depression screening scores		
PHQ-9 score (continuous)	15	15
Minimal or no depression (PHQ-9 score from 1 to 4), %	2	2
Mild depression (PHQ-9 score from 5 to 9), %	10	10
Moderate depression (PHQ-9 score from 10 to 14), %	35	36
Moderately severe depression (PHQ-9 score from 15 to 19), %	30	29
Severe depression (PHQ-9 score from 20 to 27), %	22	23
Health status and diagnoses, %		
Disabled	15	17
Select chronic conditions[a]		
Psychiatric condition	58	60
Cardiovascular disease	31	32
Pulmonary condition	23	24
Skeletal condition	20	20
Gastrointestinal condition	17	18
Diabetes	16	16
Mean CDPS score[b]	1.90	1.90
Service use during year before enrollment		
Total hospitalizations per 1,000	236	231
Total hospitalizations per 1,000, 3 months before enrollment	287	273
Any hospitalizations, %	16	15
Total ED or observation visits per 1,000	1,172	1,120
Total ED or observation visits per 1,000, 3 months before enrollment	1,512	1,423
Any ED or observation visits, %	48	49

Sources: Mathematica's analysis of information from awardee's finder file and Medicaid claims and enrollment data.

Table 3 (*continued)*

Notes: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is the date participants agreed to receive BHIP services. All beneficiary characteristics were measured during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid. Also, comparison group characteristics were weighted using inverse propensity score reweighting, which ensured the similarity between the treatment and comparison groups

None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level, two-tailed test.

Appendix B presents the full balance results.

[a] Conditions that are most common in the sample (defined as having a prevalence of greater than 10 percent) are reported.

[b] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; PHQ-9 = Patient Health Questionnaire 9-item; SE = standard error; TANF = Temporary Assistance for Needy Families.

Analytic approach

The impact estimates are based on a difference-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment relative to the change in outcomes among a comparison group. If external trends affect both groups similarly, a comparison group similar on observable and unobservable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables will capture differences on unobserved variables as well. The primary claims-based outcomes of the study are number of hospitalizations and number of ED visits. The study did not have access to reliable and complete expenditure data to evaluate the impact of the program on total cost of care because nearly all participants were in managed care programs. The study also lacked the data to evaluate the impact on clinical outcomes, such as improvement of depression symptoms.

The study defined the pre-enrollment period as the year before each participant's enrollment date and the post-enrollment period as up to the two years after, depending on when the beneficiary enrolled. The enrollment date was the date participants completed screening and agreed to participate in the BHIP. The study required assigning a pseudo-enrollment date to each comparison beneficiary, defined as the date the comparison beneficiary completed PHQ-9 screening. As mentioned previously, the comparison group was offered collocated, but not as intensive, behavioral health care at nonparticipating sites.

The intervention might have had different effects on outcomes for patients depending on the severity of their symptoms because their needs differed. Also, the BHIP's approach to treatment differed depending on severity of symptoms. For those reasons, the study separately analyzed impacts for beneficiaries with moderate or mild depression (defined as PHQ-9 scores from 5 to 14) and those with moderately severe or severe depression (PHQ-9 scores of 15 or higher). Appendix A describes the statistical models and outcomes used to estimate the effects of the

program, as well as the identification of the final analytic sample. Appendix A provides additional details on the model and outcomes used to assess the impact of the BHIP.

IMPACT RESULTS

The program had a discernible favorable impact on hospitalizations and ED visits for the full study sample in the second follow-up year (Year 2). The estimated reduction in hospitalizations for the full sample in Year 2 was relatively large (18 percent), but due to small sample sizes, it was not statistically significant. Also in Year 2, there was a 14 percent statistically significant estimated reduction in ED visits for the full sample. The estimated effect on hospitalizations was concentrated among participants with moderately severe or severe depression at enrollment (Table 4). Conversely, findings for ED visits were concentrated among participants with mild or moderate depression. Appendix C presents the full results of the impact analysis.

Even though only one result in Table 4 is statistically significant (the 14 percent reduction in ED visits in Year 2), other Year 2 results could be interpreted as meaningful because they are moderate or large in magnitude and the results show a sensible pattern of improvement. For example, impact estimates became more favorable in each quarter (data not shown). Appendix D provides results from a Bayesian analysis.

Table 4. Estimated impact of BHIP on selected outcomes

	Full group		Patients with mild or moderate depression at enrollment		Patients with moderately severe or severe depression at enrollment	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Number of hospitalizations (per 1,000 beneficiaries)						
Impact (rate)	-17	-56	27	-18	-59	-91
Percentage impact	-6.1%	-18.0%	13%	-8.1%	-17.0%	-24.0%
p-value	0.68	0.19	0.60	0.77	0.36	0.12
Number of ED visits (per 1,000 beneficiaries)						
Impact (rate)	-71	-177*	-121	-242	-25	-117
Percentage impact	-5.8%	-14.0%	-9.7%	-20.0%	-2.0%	-9.5%
p-value	0.37	0.09	0.31	0.17	0.82	0.34
Sample size						
Treatment	2,069	1,758	990	836	1,079	922
Comparison	1,432	1,162	574	451	858	711

Sources: Mathematica's analysis of information from the awardee's enrollment database from May 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data from May 1, 2014, through August 31, 2018.

Note: Impact estimates are based on the regression-adjusted difference between the treatment and comparison group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents full impact estimates. Appendix D shows the results from the Bayesian analysis.

*Significantly different from zero at the .10 level, two-tailed test.

BHIP = Behavioral Health Integration Program; ED = emergency department.

Several factors help explain these findings. First, the awardee expected that it would take time to observe substantial effects. Both for ED visits and hospitalizations, the awardee expected small improvements in Year 1 and larger improvements in Year 2, which is precisely what the findings show.

Second, the finding that participants with moderately severe or severe depression experienced a large improvement in hospitalizations is consistent with their higher levels of hospitalization compared to participants with mild or moderate depression (Appendix C, Table C.2). In other words, participants with moderately severe or severe depression were at a higher risk of hospitalizations and were thus potentially more likely to achieve improvements in these outcomes than participants with mild or moderate depression.

Favorable findings are consistent with a carefully implemented, intensive intervention. As previously mentioned, the BHIP's behavioral health staff engaged a large proportion of participants. Those engaged participants had nearly four follow-up visits on average during the first year of program implementation and, by the end of the third year, seven follow-up visits. Further, the BHIP addressed unmet needs for those who did not improve within the first three months of treatment. Many of those beneficiaries received psychiatric consultations that included further assessment and recommendations for treatment.

The study results are consistent with the opinions of Montefiore staff about the effects of the BHIP on care delivery, access, and patients' satisfaction. Even though improvements in these measures do not necessarily translate into reductions in use, they do support the possibility of such reductions. Most surveyed Montefiore staff thought that the BHIP had positive impacts on the delivery of care, access to services, and participants' satisfaction and quality of life. For example, all respondents indicated that the program had a positive impact on care coordination, and more than 90 percent noted that the program improved quality and efficiency of care, fair provision of care, quality of life, and achievement of participants' health goals. Clinicians who responded to the survey also endorsed the BHIP's positive effects, although at slightly lower rates than nonclinicians. Further, PCPs collaborated well with the behavioral health team and found value in BHIP services.

Main findings from impact evaluation

- The BHIP resulted in an estimated reduction in hospitalizations and ED visits for the full sample in the second year of the follow-up period.
- The reduction in hospitalizations was larger among patients with moderately severe or severe depression at enrollment, whereas the reduction in ED visits was larger among patients with mild or moderate depression.
- The study findings suggest that team-based, collaborative care with continuous monitoring of patients' progress and adjusting treatment as needed is more effective at reducing unnecessary service use than collocated behavioral health services that do not involve a team-based approach or as much follow-up with patients.

CONCLUSION

The BHIP provided intensive, team-based behavioral health care services to patients in a primary care setting. Behavioral health services included short-term psychotherapy, medication management, and continuous monitoring of participants' progress and adjusting treatment based on need. The BHIP reduced hospitalizations and ED visits for the full sample during the second year of the follow-up period. The estimated reduction in hospitalizations in Year 2 was concentrated among participants with moderately severe or severe depression at enrollment, whereas the reduction in ED visits was concentrated among participants with mild or moderate depression. Although most of the estimated effects were not statistically significant, the results could be interpreted as meaningful because they were moderate or large in magnitude and because they became increasingly more favorable over time, an expected and sensible pattern of change that was consistent with the awardee's expectations.

These findings suggest that team-based, collaborative care with careful monitoring of participants' progress and subsequent adjustment of treatment is more effective in reducing unnecessary inpatient and ED services than collocated behavioral health services that do not involve a team-based approach or as much follow-up with patients. Because the BHIP also offered comparison beneficiaries at the nonparticipating sites a strong intervention (collocated behavioral health services), the favorable findings were larger than expected. The reduction in hospitalizations and ED visits would likely be larger if the comparison group had not also received behavioral health services.

Limitations of evaluation

The analysis has several limitations. First, because reliable and complete expenditure data were not available, it was not possible to evaluate impacts on total cost of care. Second, also due to lack of data, the study could not evaluate the impact on clinical outcomes such as improvement of depression symptoms. Third, there is some chance for differential selection in that the treatment group beneficiaries screened positive and agreed to participate in the BHIP, whereas the comparison group beneficiaries were those who screened positive and were referred for services to the behavioral health care team. This likely resulted in more favorable estimated impacts than if comparison beneficiaries were as likely as participants to receive behavioral health care services. Fourth, there might have been some *unobservable* differences between participants and comparison group beneficiaries that for which the analysis cannot account, which could lead to biased impact estimates. Because anxiety and alcohol use disorder screening scores were not available for the comparison group, it is possible that the comparison group differed from the treatment group in terms of overall anxiety and alcohol use disorder prevalence and severity. However, because participants and comparison beneficiaries were very similar on many other characteristics, the risk of this substantially biasing impact results is low. Also, reliable information on participants' race was not available. To mitigate any potential bias, it was ensured that the treatment and comparison groups were similar on county-level population characteristics, such as the proportion of non-White residents and county-level proportion of Hispanic residents. Finally, the impact of the BHIP on all participants could not be evaluated. For example, because the comparison sites did not screen for anxiety or alcohol use disorders,

the study could not evaluate the impact of the BHIP on participants who screened positive for those two conditions alone.

PROGRAM SUSTAINABILITY

At the end of its cooperative agreement in August 2018, Montefiore planned to sustain the BHIP at its six program sites through a new payment model and with funds from the state's Delivery System Reform Incentive Payment (DSRIP). The awardee worked with two health plans to pilot test its payment model at the program sites. The awardee hoped that the revenues from the new payment model, along with the DSRIP revenues, would be sufficient to sustain the program. The awardee also trained staff to embed the changes as a standard of care, rather than view them as a temporary effort, which was key to sustainability, in part to reduce turnover among behavioral health staff who were crucial to the program.

Montefiore's proposed payment model

Montefiore developed a bundled payment model to support the BHIP for enrollees with Medicaid, Medicare Advantage, or commercial coverage. The awardee proposed setting payments at $110 to $150 per patient per month, with the final payments amounts determined in negotiation with the payers.

Behavioral health providers would receive a payment if they had at least one contact with the patient during the month and a 15 percent bonus if they met three quality targets:

1. At least 75 percent of patients received at least one follow-up contact after initial assessment.
2. At least 50 percent of patients showed a 50 percent improvement on symptom scales by the 10th week.
3. At least 75 percent of patients with no clinical improvement by Day 70 received a consultation or case review.

In addition, the program would serve Medicare fee-for-service beneficiaries through the awardee's Next Generation accountable care organization. The awardee would pay the participating primary care sites bonuses for this population through its shared savings agreement.

During the 12-month extension period from September 2017 through August 2018, Montefiore expanded the program beyond the 6 original health centers to the remainder of its 22 health centers. The program implemented at these new sites was not as robust as the original program in two ways: (1) sites needed considerable time to tailor program elements to each site's needs, culture, and operating procedures, and these modifications typically hindered fidelity to the program; and (2) no involvement of patient educators until the awardee could support their salaries. The awardee expected that the new sites would meet the standards of the original program with more time, training, and funding.

REFERENCES

Blackmore, M.A., K.E. Carleton, S.M. Ricketts, U.B. Patel, D. Stein, A. Mallow, J.P. DeLuca, and H. Chung. "Comparison of Collaborative Care and Colocation Treatment for Patients with Clinically Significant Depression Symptoms in Primary Care." *Psychiatric Services,* vol. 69, no. 11, 2018, pp. 1184–1187.

Kronick, R., T. Gilmer, T. Dreyfus, and L. Lee. "Improving Health-Based Payment for Medicaid Beneficiaries: CDPS." *Health Care Financing Review,* vol. 21, no. 3, 2000, p. 29.

Zimmerman, M. "Using the 9-Item Patient Health Questionnaire to Screen for and Monitor Depression." *JAMA,* vol. 322, no. 21, 2019, pp. 2125–2126. Available at https://doi.org/10.1001/jama.2019.15883.

Appendix A

Description of modeling strategy and outcome variables

This page has been left blank for double-sided copying.

The impact estimates for hospitalizations and number of emergency department (ED) visits are based on a propensity score weighted difference-in-differences approach with beneficiary-level fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of core outcomes used for this evaluation.

Among 6,559 patients enrolled between February 1, 2015 and August 31, 2017, 2,450 were not Medicaid beneficiaries. The impact analysis included about half of all program participants who were Medicaid beneficiaries, according to awardee data (Table A.1). Participants were patients who completed the Patient Health Questionnaire 9-Item (PHQ-9) instrument and who agreed to participate in the program from May 1, 2015, to August 31, 2017, as reported in the awardee's database. The study dropped more than three-quarters of the excluded participates because they did not meet the standard eligibility and claims-based inclusion criteria. It did not include another 10 percent because they were younger than 19; and another 10 percent because they enrolled in the first three months of program operations. The remaining 2.5 percent were excluded because they were either missing a PHQ-9 score or their PHQ-9 score was zero.

The analysis separately analyzed impacts for beneficiaries with moderate or mild depression (defined as PHQ-9 scores from 5 to 14) and those with moderately severe or severe depression (PHQ-9 scores of 15 or higher). In creating these cutoffs, the analysis followed scoring from the literature (PHQ-9 instructions; Zimmerman 2019).

Table A.1. Identifying the final sample for impact analysis for Montefiore

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total participants through August 31, 2017		**6,559**
Participants were not Medicaid beneficiaries	2,450	4,109
Total Medicaid program participants through August 31, 2017		**4,109**
Participants did not meet standard eligibility and claims-based inclusion criteria[a]	1,574	2,535
Participants younger than 19	209	2,326
Participants enrolled in first three months of program operations[b]	206	2,120
Participants with missing PHQ-9 score	35	2,085
Participants with PHQ-9 score of zero at enrollment	16	2,069
Final analytic sample		**2,069**

Sources: Mathematica's analysis of information from the awardee's enrollment database from May 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data from May 1, 2014, through August 31, 2018.

[a] Participants were excluded from the analysis sample if they met one or more of the following exclusion criteria: (1) not enrolled in Medicaid for at least 90 eligible days in the baseline year; (2) dually eligible on their enrollment date

Table A.1 (*continued*)

(eligible for both Medicare and Medicaid) because, without Medicare claims, their outcomes could not be accurately measured; (3) were eligible for only restricted Medicaid benefits; (4) had some type of third-party coverage; (5) were enrolled in CHIP; or (6) died within 30 days of enrollment.

[b] The study excluded participants who enrolled in the first three months of program because no comparison beneficiaries enrolled during that time frame.

CHIP = Children's Health Insurance Program; PHQ-9 = Patient Health Questionnaire 9-Item.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for propensity score reweighting. The table displays the weighted means of baseline characteristics for the 2,069 treatment beneficiaries and the 1,432 comparison beneficiaries used in the impact analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable. The standardized difference was calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The reweighting variables included demographic characteristics (age, gender, and race); location; health status (as measured by the Patient Health Questionnaire 9-Item [PHQ-9] score, Chronic Illness and Disability Payment System [CDPS] score, and CDPS condition indicators); and baseline year service use. The study team separately generated propensity score weights for beneficiaries with mild or moderate depression and those with moderately severe or severe depression. The variables are measured over various specified intervals within the 12 months before enrollment in the intervention.

The table also shows the results of the equivalency-of-means tests. The *p*-values come from a weighted two-sample t-test, which provides evidence of a statistically significant difference in the means. The equivalence test *p*-values are the greater of the two one-sided weighted t-test *p*-values equivalence test, which assesses whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, an omnibus test was performed in which the null hypothesis is that the treatment and comparison groups are balanced across all linear combinations of the covariates. The results assess the closeness of fit between the treatment and comparison groups on key characteristics likely to be associated with study outcomes.

Because the study separately analyzed beneficiaries with moderate or milder depression and those with moderately severe or severe depression, it also assessed the differences between treatment and comparison beneficiaries within those subgroups. Within each of these two subgroups, treatment and comparison beneficiaries were also very similar on all measured characteristics (data not shown).

Table B.1. Baseline characteristics of treatment and comparison groups for Montefiore, weighted

Characteristic	Treatment mean (SE)	Comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	40 (0.27)	40 (0.35)	0.36 (0.41)	< +/-1	0.03	0.34	< 0.01
Age: 19 to 28, %	23 (0.92)	23 (1.1)	-0.92 (1.4)	-4.1	-0.02	0.42	< 0.01
Age: 29 to 36, %	20 (0.89)	21 (1.0)	-0.02 (1.3)	< +/-1	0.00	0.93	< 0.01
Age: 37 to 45, %	20 (0.87)	19 (1.0)	0.43 (1.3)	2.2	0.01	0.75	< 0.01
Age: 46 to 53, %	20 (0.87)	19 (1.1)	0.64 (1.2)	3.3	0.02	0.60	< 0.01
Age: 54 to 84, %	18 (0.84)	18 (1.1)	-0.13 (1.2)	< +/-1	0.00	0.96	< 0.01
Male, %	21 (0.90)	20 (1.1)	1.6 (1.3)	7.4	0.04	0.34	< 0.01
Area of residence, %							
New York City (five boroughs) and Long Island	93 (0.57)	95 (0.45)	-2.5 (0.77)	-2.7	-0.10	< 0.01	< 0.01
Outside New York City and Long Island	7.2 (0.57)	4.7 (0.45)	2.5 (0.77)	34	0.10	< 0.01	< 0.01
Medicaid enrollment, %							
Enrolled in comprehensive managed care	89 (0.70)	89 (0.82)	-0.09 (0.99)	< +/-1	0.00	0.39	< 0.01
Depression screening scores							
PHQ-9 score	15 (0.11)	15 (0.15)	-0.04 (0.17)	< +/-1	-0.01	0.74	< 0.01
Minimal or no depression (PHQ-9 score from 1 to 4), %	2.2 (0.32)	2.2 (0.46)	0.02 (0.45)	< +/-1	0.00	0.97	< 0.01
Mild depression (PHQ-9 score from 5 to 9), %	10 (0.66)	9.6 (0.87)	0.57 (0.98)	5.6	0.02	0.51	< 0.01
Moderate depression (PHQ-9 score from 10 to 14), %	35 (1.1)	36 (1.1)	-0.85 (1.5)	-2.4	-0.02	0.51	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Moderately severe depression (PHQ-9 score from 15 to 19), %	30 (1.0)	29 (1.2)	0.59 (1.5)	2.0	0.01	0.61	< 0.01
Severe depression (PHQ-9 score from 20 to 27), %	22 (0.92)	23 (1.2)	-0.33 (1.4)	-1.5	-0.01	0.76	< 0.01
Health status and diagnoses, %							
CDPS score[a]	1.9 (0.04)	1.9 (0.05)	-0.04 (0.05)	-1.9	-0.02	0.93	< 0.01
Disabled	15 (0.79)	17 (1.0)	-1.4 (1.2)	-9.5	-0.04	0.55	< 0.01
Psychiatric condition	58 (1.1)	60 (1.3)	-1.7 (1.6)	-3.0	-0.04	0.60	< 0.01
Cardiovascular disease	31 (1.0)	32 (1.2)	-0.34 (1.5)	-1.1	-0.01	0.85	< 0.01
Pulmonary condition	23 (0.92)	24 (1.1)	-1.4 (1.3)	-5.9	-0.03	0.54	< 0.01
Skeletal condition	20 (0.89)	20 (1.1)	0.10 (1.3)	< +/-1	0.00	0.67	< 0.01
Gastrointestinal condition	17 (0.82)	18 (1.00)	-0.92 (1.2)	-5.5	-0.02	0.68	< 0.01
Diabetes	16 (0.81)	16 (1.1)	-0.54 (1.2)	-3.4	-0.01	0.85	< 0.01
Quarter of enrollment, %							
Quarter 1	4.5 (0.46)	4.7 (0.70)	-0.13 (0.66)	-2.8	-0.01	0.70	< 0.01
Quarter 2	11 (0.68)	11 (0.85)	-0.05 (0.98)	< +/-1	0.00	0.67	< 0.01
Quarter 3	11 (0.69)	10 (0.77)	0.80 (1.0)	7.1	0.03	0.75	< 0.01
Quarter 4	14 (0.76)	14 (0.76)	0.11 (1.1)	< +/-1	0.00	0.87	< 0.01
Quarter 5	13 (0.75)	13 (0.82)	0.27 (1.1)	2.0	0.01	0.68	< 0.01
Quarter 6	11 (0.70)	11 (0.85)	0.16 (1.0)	1.4	0.01	0.75	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Quarter 7	11 (0.69)	11 (0.88)	-0.32 (1.0)	-2.9	-0.01	0.96	< 0.01
Quarter 8	11 (0.69)	11 (0.80)	-0.23 (1.0)	-2.0	-0.01	0.88	< 0.01
Quarter 9	8.7 (0.62)	8.3 (0.79)	0.32 (0.87)	3.7	0.01	0.58	< 0.01
Quarter 10	4.2 (0.44)	5.1 (0.65)	-0.94 (0.71)	-22	-0.05	0.25	< 0.01
Service utilization							
Total hospitalizations, per 1,000 beneficiaries	236 (16)	231 (24)	4.4 (23)	1.9	0.01	0.83	< 0.01
Total hospitalizations, per 1,000 beneficiaries, 3 months before enrollment	287 (26)	273 (43)	14 (40)	4.8	0.01	0.72	< 0.01
Any hospitalizations, %	15 (0.79)	15 (1.00)	0.35 (1.2)	2.3	0.01	0.58	< 0.01
Total ED or observation visits, per 1,000 beneficiaries	1,172 (45)	1,120 (57)	51 (59)	4.4	0.03	0.44	< 0.01
Total ED or observation visits, per 1,000 beneficiaries, 3 months before enrollment	1,512 (81)	1,423 (83)	88 (104)	5.8	0.02	0.35	< 0.01
Any ED or observation visits, %	46 (1.1)	48 (1.3)	-1.6 (1.6)	-3.6	-0.03	0.58	< 0.01
Primary care visits, per 1,000 beneficiaries, ambulatory setting	3,820 (90)	3,911 (117)	-91 (148)	-2.4	-0.02	0.62	< 0.01
Primary care visits, per 1,000 beneficiaries, ambulatory setting, 3 months before enrollment	5,870 (131)	5,862 (150)	7.3 (204)	< +/-1	0.00	0.74	< 0.01
Specialist visits, per 1,000 beneficiaries, any setting	6,405 (201)	6,513 (288)	-108 (298)	-1.7	-0.01	0.99	< 0.01
Specialist visits, per 1,000 beneficiaries, any setting, 3 months before enrollment	8,021 (269)	8,421 (423)	-399 (450)	-5.0	-0.03	0.55	< 0.01
Specialist visits, per 1,000 beneficiaries, ambulatory setting	4,990 (162)	5,163 (227)	-173 (245)	-3.5	-0.02	0.78	< 0.01
Specialist visits, per 1,000 beneficiaries, ambulatory setting, 3 months before enrollment	6,263 (219)	6,497 (297)	-234 (354)	-3.7	-0.02	0.76	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total 30-day unplanned readmissions, %	24 (5.2)	22 (7.4)	2.2 (6.8)	9.0	0.01	0.70	< 0.01
Area-level factors							
Median county household income, 2015	40,246 (321)	39,140 (287)	1,106 (439)	2.7	0.08	0.02	< 0.01
Percentage Hispanic in zip code of residence	0.56 (0.00)	0.57 (0.00)	-0.01 (0.01)	-1.8	-0.06	0.08	< 0.01
Percentage non-White in zip code of residence	0.78 (0.00)	0.78 (0.00)	-0.01 (0.00)	-1.0	-0.05	0.14	< 0.01
Propensity score	0.64 (0.00)	0.63 (0.00)	0.01 (0.00)	1.5	0.06	0.11	< 0.01
Number of beneficiaries	2069	1432					
Omnibus test[b]				Chi-squared statistic 992.44	Degrees of freedom 78.00	P-value 0.00	

Sources: Mathematica's analysis of information from the awardee's enrollment database from May 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data from May 1, 2014, through August 31, 2018. Area level factors from the Area Health Resource File (AHRF) 2016-2017 release. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable.

The study team separately generated propensity score weights for patients with mild or moderate depression symptoms (PHQ-9 scores less than 15) and those with moderately severe or severe symptoms (PHQ-9 scores of 15 or more).

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

[b] The omnibus test includes additional characteristics not shown in the table, namely: health conditions with prevalence less than 10 percent, number of days observed, proportion of adults in county with a four-year college degree, proportion of population older than 25, proportion of population older than 25 with four-year college degree, number of hospital beds in county, mental health professional shortage areas, and HCBS waivers.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; HCBS = Home and Community Based Services; PHQ-9 = Patient Health Questionnaire 9-item; SE = standard error; TANF = Temporary Assistance for Needy Families.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates

This page has been left blank for double-sided copying.

Tables C.1 and C.2 display the results from the impact analysis. Table C.1 shows the impact estimates for the full study population, measured separately over intervention Years 1 and 2 and over both years together. Table C.2 shows similar results for the subgroups of 990 treatment beneficiaries (48 percent) who had mild or moderate depression at enrollment and the 1,079 treatment beneficiaries (52 percent) who had moderately severe or severe depression at enrollment. The models were estimated over the number of hospitalizations per 1,000 beneficiaries and number of emergency department (ED) or observation visits per 1,000 beneficiaries. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value, defined as the treatment group mean minus the impact estimate. An asterisk indicates impact estimates that differ statistically from zero at the .10 level, using a two-tailed test.

Further, none of the results changed meaningfully when outliers were capped at the 98th percentile, or when a two-year baseline period was used instead of a one-year baseline period (data not shown).

Table C.1. Estimated impact of the Montefiore intervention on ED visits and inpatient admissions for all beneficiaries during one- and two-year follow-up periods

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Number of hospitalizations, per 1,000 beneficiaries					
Baseline year	236	232			
Year 1	274	287	-17 (42)	-6.1%	0.68
Year 2	254	305	-56 (42)	-18.0%	0.19
Cumulative	266	294	-32 (37)	-11%	0.38
Number of ED or observation visits, per 1,000 beneficiaries					
Baseline year	1,166	1,118			
Year 1	1,147	1,171	-71 (80)	-5.8%	0.37
Year 2	1,012	1,141	-177* (106)	-14.0%	0.09
Cumulative	1,096	1,160	-112 (78)	-9.1%	0.15
Sample sizes					
Number of beneficiaries					
Baseline year	2,069	1,432			
Year 1	2,069	1,432			
Year 2	1,758	1,162			
Cumulative	2,069	1,432			

Sources: Mathematica's analysis of information from the awardee's enrollment database from May 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data as of August 31, 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Table C.1 (*continued*)

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

ED = emergency department; SE = standard error.

Table C.2. Estimated impact of the Montefiore intervention on ED visits and inpatient admissions during one- and two-year follow-up periods, by subgroup

	Beneficiaries with mild or moderate depression at enrollment (PHQ-9 from 1 to 14)					Beneficiaries with moderately severe or severe depression at enrollment (PHQ-9 from 15 to 27)				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries										
Baseline year	213	207				258	255			
Year 1	244	211	27 (52)	13.0%	0.60	301	356	-59 (65)	-17.0%	0.36
Year 2	217	230	-18 (62)	-8.1%	0.77	288	376	-91 (58)	-24.0%	0.12
Cumulative	234	218	9.4 (48)	4.4%	0.84	296	364	-71 (57)	-19%	0.21
ED or observation visits, per 1,000 beneficiaries										
Baseline year	1,130	1,046				1,198	1,185			
Year 1	1,110	1,147	-121 (118)	-9.7%	0.31	1,181	1,193	-25 (108)	-2.0%	0.82
Year 2	940	1,098	-242 (175)	-20.0%	0.17	1,077	1,182	-117 (123)	-9.5%	0.34
Cumulative	1,044	1,127	-168 (120)	-14%	0.16	1,143	1,190	-60 (102)	-4.9%	0.56
Sample sizes										
Number of beneficiaries										
Baseline year	990	574				1,079	858			
Year 1	990	574				1,079	858			
Year 2	836	451				922	711			
Cumulative	990	574				1,079	858			

Sources: Mathematica's analysis of information from the awardee's enrollment database from May 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data as of August 31, 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

ED = emergency department; SE = standard error.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for Montefiore Medical Center (Montefiore) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to Montefiore. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on two core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for two core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for Montefiore led to a Bayesian estimate of the program's impact on hospital admissions and ED visits of approximately -5 percent, an estimated reduction of 14-16 hospitalizations per 1,000 beneficiaries and 61-72 emergency department (ED) visits per 1,000 beneficiaries, across the two program years.

Table D.1. Comparison of frequentist and Bayesian impact estimates for Montefiore in the first two years after enrollment

		Impact estimate (95 percent interval)			Percentage impacts	
Outcome	Follow-up period	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Hospital admissions	Year 1	-17 (-99, 64)	-16 (-41, 9.6)	-3%	-6%	-5%
	Year 2	-56 (-139, 27)	-14 (-42, 13)	-2%	-18%	-5%
ED visits	Year 1	-71 (-228, 86)	-72 (-181, 34)	-4%	-6%	-6%
	Year 2	-177 (-384, 31)	-61 (-168, 43)	-3%	-14%	-5%

Source: Mathematica's analysis of information from the awardee's enrollment database from May 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data from May 1, 2014 through August 31, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

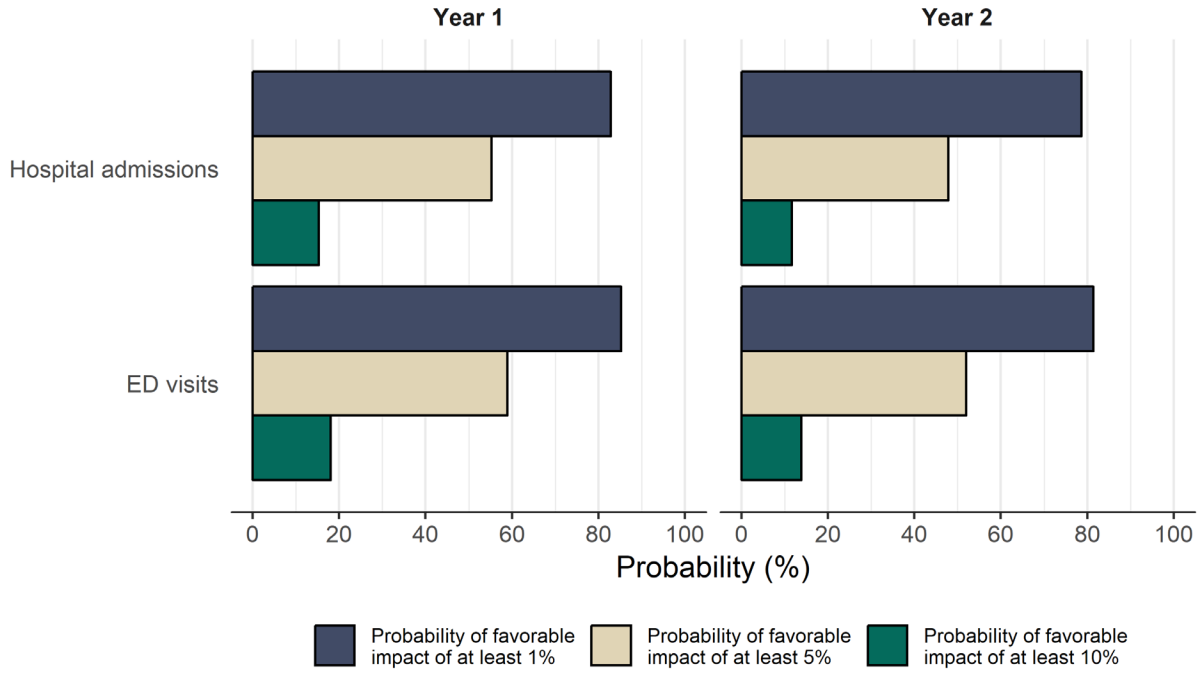
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results are somewhat imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that Montefiore achieved favorable impacts during each of the first two years on two core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the Montefiore program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from the awardee's enrollment database from May 1, 2015, through August 31, 2017, and Medicaid claims and enrollment data from May 1, 2014 through August 31, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a strong probability—in the range of 80 percent—that Montefiore had a favorable impact of 1 percent or more on hospital admissions and ED visits in each of the first two program years. The probabilities of more meaningful impacts of 5 percent or more are also moderate at 50-60 percent. Nonetheless, taken together these probabilities suggest a high probability of small effects, but a lower probability of substantial impacts. Thus, the Bayesian analysis and the frequentist analysis both suggest that the Montefiore program had favorable impacts, though the Bayesian analysis suggests that the impact estimates are likely to be smaller in size than the frequentist analysis.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: National Association of Children's Hospitals and Related Institutions

September 2020

Anna Christensen, Joe Zickafoose, Victoria Peebles, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

NATIONAL ASSOCIATION OF CHILDREN'S HOSPITALS AND RELATED INSTITUTIONS

The National Association of Children's Hospitals and Related Institutions (NACHRI) received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the Coordinating All Resources Effectively (CARE) program. The program sought to improve often disjointed and costly care for children with the most complex medical conditions, a group known as children with medical complexity (CMC). The program sought to engage children and youth enrolled in Medicaid who had lifelong or complex chronic conditions, malignancies, or catastrophic conditions. NACHRI received the grant in September 2014 and the program began enrolling children seven months later in April 2015. Care transformation activities covered under HCIA R2 ended in November 2017, after a no-cost extension. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program aimed to improve care coordination and management among CMC, improve the experience of care for these children and their caregivers, reduce family stress related to health care, and reduce Medicaid spending through a decrease in ED use and hospitalizations.
- In all, 10 participating children's hospitals and 53 affiliated primary care and complex care practices in eight states participated in a learning collaborative, implemented patient registries, identified members of the CMC's dynamic care teams, and developed access and care plans for participants and their families.
- This impact analysis relied on 3,836 Medicaid beneficiaries who received care in practices affiliated with 6 of the 10 participating hospital sites and 15,138 comparison beneficiaries with similar demographic and health characteristics who were insured by the same Medicaid payers (state or managed care organization) and lived in similar types of counties.

The awardee involved 10 children's hospitals in a learning collaborative to implement and better manage and coordinate care, engage families, and transform practice activities. Each hospital worked with one to six primary care practices and its own hospital-based program for CMC (when applicable). Through program activities, the awardee sought to improve the quality of care management and coordination processes. In improving these processes, the awardee intended for the sites to work with families to identify and address unmet needs, give clearer guidance on communicating and seeking care in the event of acute needs, offer more support for self-management, and improve communication between providers. As a result, NACHRI expected families to have better experiences and lower stress levels related to care, and improvements in the child's care would result in lower health care spending by decreasing the incidence of avoidable use, such as emergency department (ED) visits and hospitalizations. The goals of the program were to (1) improve the patients' and caregivers' experience, (2) reduce family stress related to care, and (3) reduce medical expenditures.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	NACHRI sought to achieve three primary goals: (1) improve the experience of care for CMC and their caregivers, (2) reduce family stress related to health care by 10 percent, and (3) reduce overall medical expenditures by 6.8 percent.
Major innovation	• The program conducted a learning collaborative, including primary care and hospital-based complex care practices across multiple states, to implement a package of interventions to improve care for CMC with tailoring to meet local needs. • It used claims data from multiple state Medicaid payers for participating hospital sites to identify participants, evaluate health care use, and negotiate new alternative payment models.
Program components	• **Care management and care coordination.** Provided to all participants and their caregivers through collaboration of hospital-based staff, hospital- and practice-based care coordinators, and staff in collaborating primary care practices • **Practice-based quality improvement and transformation.** Supported primary care and hospital-based complex care practices to transform care processes for CMC consistent with the principles of the medical home • **Education and training.** Conducted a learning collaborative to support changes in care based on The Breakthrough Series from the Institute for Healthcare Improvement
Target population	The program sought to engage children with CMC, defined as those classified into the 3M™ CRG software categories 5b, 6, 7, 8, or 9 by using billing or claims data. These categories encompass children with lifelong or complex chronic conditions, malignancies, and catastrophic conditions.
Participating providers	The program involved 10 children's hospitals, 44 affiliated primary care practices, and 9 affiliated hospital-based complex care practices. Of these, the analysis excluded practices affiliated with 4 children's hospitals because the payers did not supply sufficient Medicaid enrollment and claims data for children in the intervention and comparison groups. The impact analysis includes practices affiliated with 6 of the participating children's hospitals.
Total enrollment	The CARE program enrolled 8,111 children and youth from April 2015 to April 2017 (just more than 100 percent of the awardee's original enrollment goal of 8,064).
Level of engagement	The awardee reported steady increases over time in families engaging with the program, as evidenced by completion of access and care plans.
Theory of change or theory of action	NACHRI hypothesized that better care management and coordination, heightened family engagement, and practice-based quality improvement and transformation would lead to better care experiences, reduced family stress, and lower costs of providing health care to CMC.
Award amount	$23,198,916
Effective launch date	NACHRI reported that it enrolled the first participants in April 2015, and it considered the program fully implemented in May 2016.
Program settings	Hospitals, primary care practices, and hospital-based complex care practices
Market area	Urban and suburban areas in California, Colorado, District of Columbia, Florida, Missouri, Ohio, Pennsylvania, and Texas
Target outcomes	• Better care: improve patients' and caregivers' experiences • Healthier people: reduce family stress related to care by 10 percent • Smarter spending: reduce medical expenditures by 6.8 percent

Table 1 (*continued*)

Program characteristics	Description
Payment model	Each implementing site used a distinct model, including per capita care management payments, shared savings, and fee-for-service.
Sustainability plans	The awardee reported that all 10 sites continued to deliver some program services after the end of award, using a combination of internal and external funding; fee-for-service billing; and, for half of the sites, additional payment models.

CARE = Coordinating All Resources Effectively; CMC = children with medical complexity; CRG = Clinical Risk Group; NACHRI = National Association of Children's Hospitals and Related Institutions.

The CARE program enrolled 8,111 children and youth from April 2015 to April 2017 (just more than 100 percent of the awardee's original enrollment goal of 8,064), the impact analysis included only the 3,836 Medicaid beneficiaries who enrolled in the CARE program via 6 of the 10 participating sites and who had Medicaid data with sufficient quality (see Table 2 and Table A.1 for details on sample exclusions). The comparison group included 15,138 Medicaid beneficiaries with similar demographic and health characteristics who were insured by the same Medicaid payers (state or managed care organizations) but who did not receive care in participating practices. Table 2 summarizes the key features of the evaluation.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group.
Intervention group for evaluation	While the CARE program enrolled 8,111 children and youth from April 2015 to April , 2,580 children were excluded because they enrolled through 4 participating sites that could not supply Medicaid data of sufficient quality; 1,677 children were excluded because they could not be matched to the Medicaid data; and 18 children were excluded because they lacked Medicaid enrollment or claims for the entirety of the follow-up period. Therefore, the treatment group for the evaluation included 3,836 Medicaid beneficiaries who enrolled in the CARE program via 6 of the 10 participating sites (that could supply Medicaid data of sufficient quality) and who could be matched to Medicaid data.
Comparison group	The comparison group included 15,138 Medicaid beneficiaries with similar demographic and health characteristics to the treatment group and who were insured by the same Medicaid payers (state or managed care organizations) but who did not receive care in participating practices.
Limitations	(1) If the implementation and effects of the program differed in the included and excluded sites, the results of the study do not generalize to the full target population. (2) For sites in which the baseline care coordination and management practices were similar to those in the CARE program, the effects of the award would be attenuated. (3) Lack of detail about site-specific implementation differences makes it difficult to explain the varying degree of impacts across sites.

CARE = Coordinating All Resources Effectively.

PROGRAM DESIGN AND ADAPTATION

The CARE program had three components: (1) care management and care coordination, (2) practice-based quality improvement and transformation, and (3) education and training.[1]

Care management and care coordination

The CARE program focused on four care coordination and management processes—or change concepts—at participating primary care and complex care practices: (1) a patient registry, (2) shared identification of members of the child's care team (known as the dynamic care team), (3) access plans, and (4) care plans. The awardee allowed sites to determine the details of many aspects of the change concepts. Depending on the site, care management and coordination came from staff in collaborating practices; hospital-based staff, including care coordinators and social workers; or collaboration between practice and hospital-based staff. In interviews, site leaders and staff reported that they had used many of the care coordination and management processes before the award, especially in the complex care practices, but the CARE program encouraged them to be more systematic in what they were already doing and to adopt new processes to strengthen current practice.

Practice-based quality improvement and transformation

Program staff helped participating primary care and complex practices implement the change concepts through the learning collaborative described later, and the sites hired practice transformation facilitators to support local implementation. The program also encouraged sites to engage caregivers in formal program advisory roles. Local staff collected and reported back data on implementation quality measures. Practices showed steady improvement on the measures across the reporting period, although performance varied across hospital sites and practice types.

Education and training

The awardee led a learning collaborative with the participating children's hospitals and practices based on The Breakthrough Series from the Institute for Healthcare Improvement. It involved monthly virtual meetings and yearly in-person meetings that included didactic and peer-to-peer education. Over time, the collaborative added more regular virtual meetings for specific staff types, such as care coordinators and hospital executives, to provide more specific peer-to-peer learning.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

NACHRI effectively implemented the program. It reported steady increases in sites' implementation of the change concepts, levels of staffing at the sites, and practice enrollment and engagement. Site leaders and staff reported that participants' families engaged with the program.

However, in the first two years of implementation, a slow pace of enrollment, obtaining institutional review board approval from multiple sites, and difficulty obtaining Medicaid data and engaging community practices delayed the awardee's service delivery. By the final year of the cooperative agreement, the awardee had overcome these challenges.

Implications of program implementation for detecting impacts

- Site leaders reported they had used many of the care processes before the award, especially in the complex care practices, which might attenuate effects in some sites.
- Full program implementation did not occur until the final year, and the awardee did not expect to see effects of the program until after full implementation of the program.
- The awardee allowed sites to determine the details of many aspects of the change concepts, which could lead to site-level variation in implementation success and program impacts. Sites also varied by the degree of medical complexity of their participants.

ESTIMATING PROGRAM IMPACTS

Enrolling participants

The program focused on children with the most complex chronic conditions, which typically involved multiple organ systems and required the services of many medical specialists. NACHRI used the 3M™ Clinical Risk Group (CRG) claims-based algorithm to define CMC and focused on children classified in the highest CRG categories, including children with lifelong or complex chronic conditions, malignancies, and catastrophic conditions (CRG groups 5b, 6, 7, 8, and 9).

The CARE program passively enrolled eligible CMC who received care at participating complex care and primary care practices, meaning that the awardee considered them participants without having to sign up or complete any activities. The program used any of three sources to identify participants who met eligibility criteria(1) existing hospital practices for CMC, (2) analyses of administrative and billing data from the children's hospitals, and (3) referrals from providers. The intervention affected all participants indirectly through the practice-level program components, including the patient registry and activities the care coordinator did outside of direct family engagement. The intervention also affected most participants and their families directly , through actively engaging with the care coordinator and collaborating with the care team on the

access and care plans. For example, awardee metrics show that in two-thirds of sites, at least 90 percent of children had an access plan that included the program's three core elements (contingency, emergency care, and after-hours access plans).

Study sample

While the CARE program enrolled 8,111 children and youth from April 2015 to April 2017, the impact evaluation relied on 3,836 Medicaid beneficiaries who enrolled in the CARE program through 6 of the 10 participating hospital sites (for which Medicaid fee for service and Medicaid managed care payers submitted enrollment and claims data of sufficient quality), and 15,138 matched comparison beneficiaries who were not enrolled in the CARE program, who lived in similar counties and who were covered by the same Medicaid payers. Participants included in the analysis were those who received care in the primary care or complex care practices affiliated with these six sites: Children's Hospital Colorado, Children's Hospital of Philadelphia, Children's Mercy Hospital Kansas City, Cook Children's Medical Center, St. Joseph's Children's Hospital, and Wolfson Children's Hospital. The study sample does not include participants from the remaining four sites (Children's National Medical Center, Cincinnati Children's Hospital, Lucile Packard Children's Hospital Stanford, and Mattel Children's Hospital) because of the lack of data for a potential comparison group, lack of data for key measures, or other data quality issues (2,580 participants excluded). The study sample also excluded 1,695 participants from the six sites in the analysis who the Medicaid enrollment data could not identify or who lacked Medicaid claims for any portion of the follow-up period. (Appendix A, Table A.1 describes the identification of the analytic sample).

Characteristics of treatment and comparison group beneficiaries

The treatment and comparison groups in the analysis had similar characteristics at baseline (Table 3). The average age of treatment and comparison group members during the baseline year was 8. More than half the children had a CRG of 6, indicating significant chronic disease in multiple organ systems. Nearly one-third of children were in the most severely ill groups (CRG 7, 8, or 9), which include children with chronic disease in three or more organ systems; children with malignancies; and children with catastrophic conditions such as HIV, spina bifida, paralysis, or reliance on a ventilator or feeding tube. In all, 30 percent of the sample were hospitalized and nearly 60 percent had at least one ED visit in the year before enrollment in the CARE program. By design, the children in the sample were much sicker than the general child population, in which fewer than 3 percent are hospitalized in a year, and about 20 percent have an ED visit.[2] Appendix B provides the full balance results measured during the 12 months before enrollment.

[2] Witt, W.P., A.J. Weiss, and A. Elixhauser. "Overview of Hospital Stays for Children in the United States, 2012." HCUP Statistical Brief No. 187. December 2014. Available at https://www.hcup-us.ahrq.gov/reports/statbriefs/sb187-Hospital-Stays-Children-2012.jsp. Accessed January 8, 2020.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Measure	Treatment (N = 3,836)	Comparison (N = 15,138)
Demographics		
Age at enrollment, years	8	8
Age group, %		
Birth to 2 years	20	21
3 to 5 years	18	16
6 to 9 years	24	25
10 to 12 years	17	15
13 to 17 years	18	18
18 to 21 years	4	4
Female, %	44	43
Health status		
3M CRG,[a] %		
CRG 5b, Single dominant or moderate chronic disease	17	16
CRG 6, Significant chronic disease in multiple organ systems	53	54
CRGs 7 through 9	30	30
CRG 7, Dominant chronic disease in three or more organ systems	5	4
CRG 8, Dominant, metastatic, and complicated malignancies	1	5
CRG 9, Catastrophic conditions	24	21
Service use during the year before enrollment		
Any hospitalizations, %	30	29
Any ED visits, %	59	57
Number of hospital admissions (per 1,000 beneficiaries)	593	529
Number of ED visits (per 1,000 beneficiaries)	1,453	1,300
Number of ambulatory care visits (per 1,000 beneficiaries)	8,710	8,601

Source: Mathematica's analysis of information from Medicaid claims and enrollment data from April 2013 to November 2018.

Notes: The baseline year is the 365 days before each beneficiary's enrollment date. The enrollment date is the date the program passively enrolled a participant. All beneficiary characteristics were measured during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid. In addition to the number of months enrolled in Medicaid, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

Appendix B presents the full balance results. The analysis required an exact match for CRG category.

[a] 3M™ CRG scores were calculated using all claims available for each beneficiary.

CRG = Clinical Risk Group; ED = emergency department.

Data Resource Center for Child and Adolescent Health. "Interactive Data Query of the 2017-2018 National Survey of Children's Health, Indicator 4.7: Hospital Emergency Room Visit." n.d. Available at https://www.childhealthdata.org/browse/survey/results?q=7091&r=1. Accessed January 8, 2020.

Analytic approach

The impact estimates relied on a difference-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment, relative to the change in outcomes among a comparison group with similar characteristics over the same period. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables will capture differences on unobserved variables as well. The primary outcomes are number of hospital admissions and number of ED visits. A secondary outcome is the number of ambulatory care visits (primary and specialty care combined). Data on Medicaid expenditures were unavailable for this study.

The pre-enrollment period was the year before each participant's enrollment date, and the post-enrollment period was the two years after. Appendix A describes the statistical models used to estimate the effects of the program, and it identifies the final analytic sample.

IMPACT RESULTS

For the CARE participants and the matched comparison group, hospitalizations, ED use, and ambulatory care visit rates were high during the baseline year and decreased steadily in the two-year follow-up period (Table 4). Hospitalizations decreased an estimated 15 percent more among the comparison group than among CARE participants, and ambulatory care visits decreased an estimated 9 percent more among the comparison group. These differences were statistically significant, and effect sizes were larger among the sickest group of children (CRGs 7, 8, and 9). In contrast, ED use decreased more among CARE participants than among the comparison group. The program's effects on ED use grew the longer the children participated, with the reduction in the second enrollment year estimated to be 10 percent and statically significant. A subgroup analysis by the participants' level of medical complexity (CRGs 5b, 6, 7, 8, and 9) showed that the program's effect on ED use in the second year of enrollment were larger and more favorable among the sickest group of children (CRGs 7, 8, and 9), with an estimated 15 percent reduction attributable to the CARE program.

The effects of the CARE program varied across the six sites included in the analysis. Estimates of the program's impact on hospitalization rates in the second year of enrollment ranged widely across sites (from a 35 percent reduction to a 55 percent increase) (Table 5). The program's effects on ED visits, however, were favorable across all sites. The estimated effects for each site ranged from a 3 percent reduction to a 22 percent reduction in ED visits, although not all estimates were statistically significant. Appendix C presents the full results of this analysis. Appendix D presents the results from the Bayesian analysis.

Table 4. Estimated impact of CARE program on outcomes, by year and cumulatively

	Baseline	Follow-up Year 1	Follow-up Year 2	Cumulative follow-up
Number of hospitalizations, per 1,000 beneficiaries				
Treatment group mean	593	456	379	419
Comparison group mean	529	329	264	299
Impact (rate)		63*	51*	56*
Percentage impact		16.0%	16.0%	15.0%
p-value		0.01	0.07	0.02
Number of ED visits, per 1,000 beneficiaries				
Treatment group mean	1,453	1,246	1,056	1,161
Comparison group mean	1,300	1,118	1,023	1,077
Impact (rate)		-25	-120*	-68*
Percentage impact		-1.9%	-10.0%	-5.5%
p-value		0.53	< 0.01	0.06
Number of ambulatory care visits, per 1,000 beneficiaries				
Treatment group mean	8,710	7,406	6,258	6,901
Comparison group mean	8,601	6,787	5,453	6,204
Impact (rate)		510*	698*	588*
Percentage impact		7.2%	12.0%	9.1%
p-value		< 0.01	< 0.01	< 0.01
Sample size				
Treatment	3,836	3,834	3,528	3,836
Comparison	15,138	15,092	13,625	15,138

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Notes: Impact estimates relied on the regression-adjusted difference between the treatment and matched control group members. Percentage impacts were then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents full impact estimates. Appendix D presents the results from the Bayesian analysis.

* Significantly different from zero at the .10 level, two-tailed test.

CARE = Coordinating All Resources Effectively; ED = emergency department.

Table 5. Estimated Year 2 impacts of CARE program on outcomes, by site

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Number of hospitalizations, per 1,000 beneficiaries						
Impact (rate)	24	180*	111	4	52	-21
Percentage impact	8.7%	55.0%	23.0%	1.0%	11.0%	-35.0%
p-value	0.57	0.03	0.34	0.97	0.52	0.48
Number of ED visits, per 1,000 beneficiaries						
Impact (rate)	-66	-221*	-260*	-38	-189*	-100
Percentage impact	-5.3%	-22.0%	-21.0%	-2.9%	-14.0%	-15.0%
p-value	0.37	0.04	0.08	0.80	0.07	0.27

Table 5 (*continued*)

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Number of ambulatory care visits, per 1,000 beneficiaries						
Impact (rate)	1,112*	836*	-21	578*	263	-92
Percentage impact	20%	15.0%	< 1.0%	13.0%	3.6%	-2.1%
p-value	< 0.01	0.05	0.97	0.06	0.43	0.73
Sample size						
Treatment	1,411	455	267	352	655	388
Comparison	5,159	1,920	1,085	1,266	2,546	1,649

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Notes: Impact estimates are based on the regression-adjusted difference between the treatment and matched control group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents full impact estimates. Data sharing agreements stipulate that the site-specific results must not identify hospitals.

*Significantly different from zero at the .10 level, two-tailed test.

CARE = Coordinating All Resources Effectively; ED = emergency department.

The CARE program's favorable effect on ED visits is consistent with the program's focus on developing access and care plans for participants and their families. Sites that met the program's targets for access and care plans reduced ED visits more than sites that fell short of the targets (ED visit reductions of 14 to 22 percent compared with 3 to 5 percent; results not reported). The access and care plans sought to give the families clearer plans to communicate and seek care in the event of acute medical needs and to offer more support for families to self-manage the children's conditions. Several CARE staff reported the program had provided families better access to clinical advice and had emphasized teaching families what to do in response to specific concerns. They also reported practice sites made it easier for families to directly contact a member of their care team, most commonly their care coordinator. The access and care plans and direct access to care coordinators could have helped families address concerns in other ways and avoid a trip to the ED. That the ambulatory care visit rate did not decrease as much in the CARE group as in the matched comparison group might indicate a substitution effect, whereby families used ambulatory visits in place of some ED visits.

The greater reduction in ED visits in the children's second year of the follow-up also aligns with what program staff reported about implementing the change concepts and the effect on participants. The awardee expected it would take 6 to 12 months of enrollment for the program to affect participants' outcomes, which is consistent with seeing stronger effects in the children's second year compared to first year of enrollment. In addition, sites enrolled participants in the program from April 2015 to April 2017, but the awardee and site leaders did not consider the program fully implemented and operational until May 2016, halfway through enrollment. Children who enrolled early, therefore, might not have received all components of the program until their second year.

Hospitalizations decreased over time among children enrolled in the program, but this cannot be attributed to the effect of the program because hospitalizations decreased even more in the

matched comparison group. As a result, the impact estimates do not show a favorable effect of the program on hospitalizations. The awardee reported that the data from the hospital sites suggested CARE participants had fewer hospitalizations and shorter lengths of stay in the follow-up period than the preceding period; the hospitals did not, however, compare these reductions to a matched comparison group, as this analysis does. These results are a reminder that health care use among CMC varies over time, and previous studies have demonstrated that their health care use often decreases after a period of high use, even in the absence of an intervention.[3]

Main findings from impact evaluation

- The CARE program reduced ED visits among participants, with a greater reduction in the second year after enrollment.
- Effects on ED use were favorable across all six sites, although some had larger effect sizes than others, and not all estimates were statistically significant. There were larger impacts on children with the most complex medical conditions (CRGs 7, 8, and 9).
- Although hospital use decreased in the follow-up period for CARE participants, it decreased even more among the matched comparison group. This mean that the observed reductions in hospitalizations cannot be attributed to the program. As a result, the program did not have an impact on hospital use.

Unobservable differences between the groups could explain the estimates that hospital use decreased more among the comparison group than the CARE group. Although the intervention and comparison groups were well matched on CRG category, baseline health care use, demographics, and county-level characteristics, it is possible that the CARE group had more complex needs than the comparison group in ways that Medicaid claims data could not measure. The CARE program passively enrolled CMC who previously received care from the hospital-based complex care practices. It is possible that families whose children had the most complex needs sought care at those hospital-based practices, leading to adverse selection bias. In this group, hospitalizations might have been largely unavoidable, and CARE program activities might not have been able to affect them.

According to the program's theory of action, by engaging with primary care and hospital-based complex care practices, the awardee sought to increase the frequency and improve the quality of care management and coordination. As a result, families would have better experiences with care and lower stress levels related to care. Improving children's care would result in lower health care spending by decreasing the incidence of avoidable use, such as ED visits and hospitalizations. The favorable impacts of the program on ED visits is consistent with the theory of action. But because the program did not reduce hospitalizations or ambulatory care visits more than the comparison group did, it is unlikely that the program caused lower health care spending.

[3] Peltz, A., M. Hall, D.M. Rubin, K.D. Mandl, J. Neff, M. Brittan, E. Cohen, D.E. Hall, D.Z. Kuo, R. Agrawal, and J.G. Berry. "Hospital Utilization Among Children with the Highest Annual Inpatient Cost." *Pediatrics,* vol. 137, no. 2, 2016.

The analysis could not directly measure effects on Medicaid spending because many participants in the program received coverage from managed care organizations (MCOs), in which the state Medicaid agency pays the MCO a set amount per member, regardless of the services provided, and the MCOs do not report their payment rates to clinicians.

CONCLUSION

Overall, the NACHRI implemented the program successfully and it had a favorable estimated effect on one of the two primary outcomes measured for this study: ED use. It did not have a favorable effect, however, on hospitalizations. The program's effect on ED use was strongest in the second year of enrollment, and among children with the most complex chronic conditions, who had the highest rate of ED use in the baseline. Although the analysis could not directly measure Medicaid spending, it is unlikely the reduced ED use would offset the unfavorable hospitalization results to create an overall reduction in spending. Yet the estimated reduction in ED use could have improved families' experiences with care and lowered stress levels related to care, which were other goals of the program.

Limitations of evaluation

The analysis has several limitations. First, because of data availability and quality, it calculated program impacts among participants who received care in practices affiliated with only 6 of the 10 participating hospitals. If the implementation and effects of the program differed in the included and excluded sites, the results of the study do not generalize to the full target population. In addition, Medicaid enrollment and claims data could identify only about 70 percent of participants in the included sites (Appendix A, Table A.1). If there were systematic differences in participants who could versus could not be identified, the results would not generalize to that full population. Second, site leaders and program staff reported they had used many of the care processes before the award, especially in the hospital-based complex care practices. Sites varied in the degree to which they had to implement new processes, rather than change or standardize previous processes. For sites in which the baseline care coordination and management practices were similar to those in the CARE program, the effects of the award would be attenuated. Finally, lack of detail about site-specific implementation differences makes it difficult to identify site-specific characteristics that can explain the varying degree of impacts on ED visits.

PROGRAM SUSTAINABILITY

Before the end of the awardee's no-cost extension, NACHRI concluded operating the CARE learning collaborative for its 10 sites but used the no-cost extension to help sites plan to sustain aspects of their individual CARE programs. The awardee reported that all 10 sites continued to deliver some level of services through a combination of internal and external funding; fee-for-service billing; and, for half of the sites, additional payment models. Although NACHRI did not have explicit plans to scale up its program, the awardee disseminated lessons learned from its experience in hopes of generating interest among other hospitals.

NACHRI's proposed payment models

Each of the 10 sites participating in NACHRI's CARE program developed a payment model approach specific to its state and local context. Three sites implemented a per beneficiary per month care coordination fee (fee amounts varied significantly). One site established a shared savings arrangement with a provider network and the MCO, and another received direct funding from an MCO to support care coordinator positions.

By the end of the award, 5 of the 10 sites had implemented payment models through contracts with a state Medicaid agency or a Medicaid MCO. Two other sites continued to work on payment model proposals with an actuary, and another site was in negotiations with a Medicaid payer. The remaining two sites were unable to gain traction with their payment model, which they attributed primarily to lack of interest from the leaders of the health system for a model that affects a relatively small patient population.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for number of visits or stays relied on a difference-in-differences approach with beneficiary-level fixed effects. The estimates show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or emergency department visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or emergency department visit during the baseline period. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of core outcomes used for this evaluation.

The impact analysis included only 47 percent of total participants, according to awardee data (Table A.1). It defined participants as those having been passively or actively enrolled in the CARE program from April 2015 to April 2017, as reported in the awardee's final Finder File. The study dropped most of the excluded participants because they were enrolled at one of four hospitals for which the study did not have sufficient Medicaid enrollment and claims data from the participating payers (32 percent of participants). Another 21 percent of participants could not be found in the Medicaid enrollment data via their Medicaid ID or Social Security number. The remaining enrollees were dropped because they lacked any enrollment of claims data during the entire follow-up period (less than 1 percent).

Table A.1. Identifying the final sample for impact analysis for NACHRI

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through August 31, 2017		8,111
Participants enrolled in four hospitals not included in impact analysis because of Medicaid data limitations[a]	2,580	5,531
Participants who could not be identified in the Medicaid enrollment or claims files[b]	1,677	3,854
Participants who lacked Medicaid enrollment and claims data during entire follow-up period	18	3,836
Final analytic sample		3,836

Sources: Mathematica's analysis of awardee-provided data and Medicaid claims and enrollment data from April 2013 to November 2018.

[a] The study sample did not include four hospitals. The study dropped Lucile Packard Children's Hospital Stanford and Mattel Children's Hospital because the awardee did not process the California Medicaid data. It did not include Children's National Medical Center because the payer did not provide Medicaid data for a comparison group and Cincinnati Children's Hospital because the Medicaid data supplied by the payers had significant data quality issues.

[b] The awardee linked the finder file to the Medicaid enrollment and claims files, but not all participants could be identified in the Medicaid data.

NACHRI = National Association of Children's Hospitals and Related Institutions.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. The table displays the weighted means of baseline characteristics for the 3,854 treatment beneficiaries and the 15,380 matched comparison beneficiaries used in the matching analysis, although in the modeling phase the analysis dropped 18 treatment beneficiaries and 242 comparison beneficiaries. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable. The analysis calculated the standardized difference as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The analysis performed matching separately by hospital site, and each hospital enrolled participants from one Medicaid payer (that is, either the state's fee-for-service program or one Medicaid managed care organization), so there was effectively an exact match by site. Within sites, the matching variables included demographic characteristics (age, sex, and race); health status (as measured by the 3M^{TM} Clinical Risk Group [CRG] and grouped into CRG categories 5b, 6, and 7, 8, 9); service use in the 24 months before enrollment; and area-level sociodemographic characteristics (as measured in the Area Health Resources File). The analysis required an exact match for CRG category. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

The table also shows the results of the equivalency-of-means tests. The *p*-values come from a weighted two-sample t-test, which provides evidence of a statistically significant difference in the means. The equivalence test *p*-values are the greater of the two one-sided weighted t-test *p*-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, an omnibus test was performed in which the null hypothesis is that the treatment and matched comparison groups are balanced across all linear combinations of the covariates. The analysis used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely associated with study outcomes.

Table B.1. Baseline characteristics of all beneficiaries in treatment and matched comparison groups for NACHRI

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	7.9 (0.08)	8.0 (0.04)	-0.13 (0.13)	-1.6	-0.02	0.30	< 0.01
Female, %	43 (0.80)	43 (0.40)	0.07 (1.1)	< +/-1	0.00	0.95	< 0.01
White, %	21 (0.66)	21 (0.33)	0.08 (0.88)	< +/-1	0.00	0.93	< 0.01
Black, %	8.6 (0.45)	7.3 (0.20)	1.3 (0.62)	15	0.05	0.03	< 0.01
American Indian or Alaska Native, %	0.10 (0.05)	0.04 (0.01)	0.06 (0.06)	60	0.02	0.29	< 0.01
Hispanic, %	7.5 (0.42)	7.3 (0.22)	0.18 (0.64)	2.4	0.01	0.78	< 0.01
Asian, %	4.1 (0.32)	5.0 (0.18)	-0.91 (0.47)	-22	-0.05	0.05	< 0.01
Native Hawaiian or other Pacific Islander, %	2.3 (0.24)	2.5 (0.12)	-0.20 (0.35)	-9.0	-0.01	0.55	< 0.01
Other, %	29 (0.73)	29 (0.36)	-0.33 (1.0)	-1.1	-0.01	0.75	< 0.01
Missing or unknown, %	28 (0.72)	28 (0.37)	-0.20 (0.96)	< +/-1	0.00	0.84	< 0.01
Medicaid eligibility							
Months Medicaid eligible in baseline	12 (0.03)	12 (0.02)	-0.05 (0.05)	< +/-1	-0.02	0.33	< 0.01
Service use							
Total hospitalizations, 12 to 24 months before enrollment	0.51 (0.02)	0.45 (0.01)	0.06 (0.03)	12	0.05	0.02	< 0.01
Total hospitalizations, 12 months before enrollment	0.57 (0.02)	0.51 (0.01)	0.06 (0.03)	10	0.05	0.04	< 0.01
Total hospitalizations, 3 months before enrollment	0.14 (0.01)	0.14 (0.00)	0.00 (0.01)	-3.4	-0.01	0.67	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total hospital days, 12 to 24 months before enrollment	2.4 (0.20)	2.0 (0.06)	0.46 (0.25)	19	0.04	0.07	< 0.01
Total hospital days, 12 months before enrollment	2.9 (0.21)	2.2 (0.05)	0.67 (0.26)	23	0.05	< 0.01	< 0.01
Total hospital days, 3 months before enrollment	0.76 (0.07)	0.64 (0.02)	0.12 (0.09)	16	0.03	0.20	< 0.01
Total ED visits, 12 to 24 months before enrollment	1.2 (0.03)	1.1 (0.01)	0.07 (0.04)	5.8	0.04	0.10	< 0.01
Total ED visits, 12 months before enrollment	1.4 (0.03)	1.2 (0.01)	0.14 (0.04)	10	0.07	< 0.01	< 0.01
Total ED visits, 3 months before enrollment	0.34 (0.01)	0.31 (0.00)	0.03 (0.02)	8.0	0.04	0.10	< 0.01
Total outpatient visits, 12 to 24 months before enrollment	7.4 (0.11)	7.3 (0.05)	0.07 (0.17)	< +/-1	0.01	0.68	< 0.01
Total outpatient visits, 12 months before enrollment	8.3 (0.11)	8.2 (0.05)	0.10 (0.17)	1.2	0.01	0.57	< 0.01
Total outpatient visits, 3 months before enrollment	2.1 (0.04)	2.1 (0.02)	-0.01 (0.05)	< +/-1	-0.01	0.82	< 0.01
Area-level factors							
No part of county of residence designated HPSA, %	3.7 (0.30)	3.7 (0.15)	-0.06 (0.62)	-1.7	0.00	0.92	< 0.01
Entire county of residence designated HPSA, %	0.42 (0.10)	0.56 (0.06)	-0.14 (0.26)	-33	-0.02	0.60	< 0.01
One or more parts of county of residence designated HPSA, %	96 (0.32)	96 (0.16)	0.17 (0.65)	< +/-1	0.01	0.80	< 0.01
Median household income	58,395 (185)	58,619 (97)	-223 (269)	< +/-1	-0.02	0.41	< 0.01
Metropolitan area of 1 million people, %	94 (0.40)	91 (0.23)	2.3 (0.68)	2.4	0.09	< 0.01	< 0.01
Metropolitan area with 250,000 to 1 million people, %	5.5 (0.37)	7.6 (0.21)	-2.1 (0.64)	-38	-0.09	< 0.01	< 0.01
Metro area with population fewer than 250,000, %	0.47 (0.11)	0.68 (0.07)	-0.21 (0.22)	-44	-0.03	0.36	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Urban area with population 20,000 or more, adjacent to metro area, %	0.21 (0.07)	0.12 (0.03)	0.09 (0.15)	45	0.02	0.53	< 0.01
Urban area with population 2,500 to 19,999, adjacent to metro area, %	0.19 (0.07)	0.10 (0.03)	0.09 (0.12)	49	0.02	0.43	< 0.01
Urban area with population 2,500 to 19,999, not adjacent to metro area, %	0.56 (0.12)	0.56 (0.06)	0.00 (0.27)	< +/-1	0.00	1.00	< 0.01
Rural area or fewer than 2,500 urban population, adjacent to metro area, %	0.21 (0.07)	0.31 (0.04)	-0.10 (0.19)	-48	-0.02	0.60	< 0.01
Rural area or fewer than 2,500 urban population, not adjacent to metro area, %	0.07 (0.04)	0.17 (0.03)	-0.10 (0.13)	-140	-0.04	0.45	< 0.01
Percentage of adults 25 or older in the county with a four-year degree	33 (0.13)	33 (0.07)	-0.03 (0.20)	< +/-1	0.00	0.89	< 0.01
Percentage White, 2010	69 (0.20)	69 (0.10)	-0.83 (0.30)	-1.2	-0.07	< 0.01	< 0.01
Number of general and pediatric hospital beds in county	50 (0.89)	48 (0.45)	1.8 (1.3)	3.5	0.03	0.17	< 0.01
Propensity score	0.02 (0.00)	0.02 (0.00)	0.00 (0.00)	17	0.09	< 0.01	< 0.01
Number of beneficiaries	3,854	15,380					
Omnibus test				Chi-squared statistic 1106.31	Degrees of freedom 65.00	*p*-value 0.00	

Source: Mathematica's analysis of information from Medicaid claims and enrollment data from April 2013 to November 2018. Area-level factors from the Area Health Resource File (AHRF) 2016-2017 release.

Notes: Standard errors are in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid. The analysis required an exact match for CRG category.

CRG = Clinical Risk Group; ED = emergency department; HPSA = health professionals shortage area; NACHRI = National Association of Children's Hospitals and Related Institutions; SE = standard error.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Tables C.1 through C.6 display the results from the impact analysis. Table C.1 shows the impact estimates for the full study population, measured separately over intervention Years 1 and 2, as well as cumulative results. Tables C.2 and C.3 show similar results by Clinical Risk Group (CRG) subgroup (CRG 5b, 6, 7, 8, 9), and Tables C.4 through C.6 show similar results by hospital site. The impact analysis estimated models over number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that statistically differ from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the NACHRI intervention on select Medicaid use measures during a 24-month follow-up period

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries					
Baseline year	593	529			
Months 1-12	456	329	63** (26)	16%	0.01
Months 13-24	379	264	51* (28)	16%	0.07
Months 1-24	419	299	56** (24)	15%	0.02
Hospital days per beneficiary (log transformed)					
Baseline year	0.60	0.54			
Months 1-12	0.48	0.33	0.09*** (0.03)	22%	< 0.01
Months 13-24	0.37	0.27	0.05* (0.03)	15%	0.09
Months 1-24	0.66	0.47	0.13*** (0.03)	25%	< 0.01
ED visits, per 1,000 beneficiaries					
Baseline year	1,453	1,300			
Months 1-12	1,246	1,118	-25 (39)	-1.9%	0.53
Months 13-24	1,056	1,023	-120*** (44)	-10%	< 0.01
Months 1-24	1,161	1,077	-68* (37)	-5.5%	0.06
Outpatient visits, per 1,000 discharges					
Baseline year	8,710	8,601			
Months 1-12	7,406	6,787	510*** (116)	7.2%	< 0.01
Months 13-24	6,258	5,453	696*** (135)	12%	< 0.01
Months 1-24	6,901	6,204	588*** (113)	9.1%	< 0.01
Percentage with any hospitalization					
Baseline year	30	29			
Months 1-12	24	17	7.2*** (0.76)	43%	< 0.01
Months 13-24	19	14	5.1*** (0.77)	37%	< 0.01
Percentage with any ED visit					
Baseline year	59	57			
Months 1-12	53	51	2.6*** (0.91)	5.1%	< 0.01
Months 13-24	48	46	1.8* (0.98)	3.9%	0.06

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes					
Number of beneficiaries					
Baseline year	3,836	15,138			
Months 1-12	3,834	15,092			
Months 13-24	3,528	13,625			
Months 1-24	3,836	15,138			

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Notes: Impact estimates for number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; NACHRI = National Association of Children's Hospitals and Related Institutions; SE = standard error.

Table C.2. Estimated impact of the NACHRI intervention on select Medicaid use measures during a 24-month follow-up period, by 3M CRG

	Beneficiaries with CRG 5b: Significant lifelong chronic diseases				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries					
Baseline year	244	221			
Months 1-12	221	138	59 (39)	34%	0.13
Months 13-24	171	122	25 (43)	15%	0.56
Months 1-24	196	130	42 (35)	25%	0.23
Hospital days per beneficiary (log transformed)					
Baseline year	0.28	0.26			
Months 1-12	0.25	0.16	0.07 (0.05)	36%	0.12
Months 13-24	0.19	0.11	0.06 (0.05)	39%	0.23
Months 1-24	0.39	0.24	0.13*** (0.05)	46%	< 0.01
ED visits, per 1,000 beneficiaries					
Baseline year	1,210	947			
Months 1-12	1,013	803	-54 (78)	-5.0%	0.49
Months 13-24	884	703	-82 (88)	-8.2%	0.35
Months 1-24	954	760	-69 (73)	-6.6%	0.34

Table C.2 (*continued*)

	Beneficiaries with CRG 5b: Significant lifelong chronic diseases				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Outpatient visits, per 1,000 discharges					
Baseline year	5,767	5,581			
Months 1-12	4,564	4,311	68 (193)	1.5%	0.72
Months 13-24	4,059	3,548	326 (241)	8.5%	0.18
Months 1-24	4,341	3,965	191 (193)	4.5%	0.32
Percentage with any hospitalization					
Baseline year	15	15			
Months 1-12	14	10	4.1*** (1.5)	41%	< 0.01
Months 13-24	12	8.1	4.2*** (1.6)	51%	< 0.01
Percentage with any ED visit					
Baseline year	52	51			
Months 1-12	48	44	4.6** (2.2)	11%	0.03
Months 13-24	45	39	5.4** (2.4)	14%	0.02
Sample sizes					
Number of beneficiaries					
Baseline year	634	2,719			
Months 1-12	634	2,709			
Months 13-24	579	2,420			
Months 1-24	634	2,719			

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Note: Impact estimates for number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

CRG = Clinical Risk Group; ED = emergency department; NACHRI = National Association of Children's Hospitals and Related Institutions; SE = standard error.

Table C.3. Estimated impact of the NACHRI intervention on select Medicaid use measures during a 24-month follow-up period, by 3M CRG

	Beneficiaries with CRG 6: Significant chronic diseases in multiple organ systems					Beneficiaries with CRGs 7, 8, and 9: Dominant chronic diseases in three or more organ systems, metastatic malignancy, and catastrophic conditions				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries										
Baseline year	491	428				962	874			
Months 1-12	302	206	33 (29)	11%	0.26	853	648	117* (63)	16%	0.06
Months 13-24	237	164	11 (32)	5.0%	0.74	739	516	136* (70)	24%	0.05
Months 1-24	271	186	23 (28)	8.9%	0.42	798	591	120** (60)	18%	0.05
Hospital days per beneficiary (log transformed)										
Baseline year	0.52	0.45				0.95	0.90			
Months 1-12	0.34	0.24	0.03 (0.03)	9.0%	0.35	0.87	0.62	0.20**** (0.06)	31%	< 0.01
Months 13-24	0.24	0.18	-0.01 (0.03)	-5.4%	0.70	0.73	0.54	0.15** (0.07)	27%	0.03
Months 1-24	0.47	0.34	0.06* (0.03)	14%	0.07	1.2	0.87	0.27*** (0.06)	30%	< 0.01
ED visits, per 1,000 beneficiaries										
Baseline year	1,521	1,431				1,464	1,260			
Months 1-12	1,284	1,210	-16 (55)	-1.2%	0.77	1,307	1,127	-25 (75)	-1.8%	0.74
Months 13-24	1,090	1,085	-86 (60)	-7.2%	0.15	1,091	1,089	-203** (88)	-15%	0.02
Months 1-24	1,197	1,154	-47 (51)	-3.8%	0.36	1,211	1,113	-106 (71)	-7.9%	0.14
Outpatient visits, per 1,000 discharges										
Baseline year	8,569	8,282				10,555	10,795			
Months 1-12	7,125	6,393	446*** (143)	6.5%	< 0.01	9,438	8,822	856*** (266)	9.9%	< 0.01
Months 13-24	5,926	5,081	559*** (168)	10%	< 0.01	8,035	7,140	1,135*** (304)	16%	< 0.01
Months 1-24	6,593	5,805	502*** (141)	8.0%	< 0.01	8,828	8,123	947*** (256)	12%	< 0.01
Percentage with any hospitalization										
Baseline year	27	26				42	41			
Months 1-12	19	13	5.8*** (0.95)	44%	< 0.01	39	28	12*** (1.7)	42%	< 0.01
Months 13-24	13	9.7	3.5*** (0.93)	36%	< 0.01	32	23	8.4*** (1.7)	36%	< 0.01
Percentage with any ED visit										
Baseline year	60	59				60	57			
Months 1-12	55	52	2.6** (1.2)	5.0%	0.03	54	52	1.6 (1.7)	3.0%	0.37
Months 13-24	48	48	0.56 (1.3)	1.2%	0.67	49	47	2.2 (1.9)	4.6%	0.25

Table C.3 (*continued*)

	Beneficiaries with CRG 6: Significant chronic diseases in multiple organ systems					Beneficiaries with CRGs 7, 8, and 9: Dominant chronic diseases in three or more organ systems, metastatic malignancy, and catastrophic conditions				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes										
Number of beneficiaries										
Baseline year	2,048	8,689				1,154	3,730			
Months 1-12	2,048	8,662				1,152	3,721			
Months 13-24	1,871	7,842				1,078	3,363			
Months 1-24	2,048	8,689				1,154	3,730			

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Note: Impact estimates for number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

CRG = Clinical Risk Group; ED = emergency department; NACHRI = National Association of Children's Hospitals and Related Institutions; SE = standard error.

Table C.4. Estimated impact of the NACHRI intervention on select Medicaid use measures during a 24-month follow-up period, by hospital site

	Site 1					Site 2				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries										
Baseline year	516	431				679	628			
Months 1-12	366	262	20 (39)	5.6%	0.61	632	333	248**** (64)	64%	< 0.01
Months 13-24	301	193	24 (42)	8.7%	0.57	499	267	180** (83)	55%	0.03
Months 1-24	333	227	21 (37)	6.6%	0.57	590	308	230*** (60)	64%	< 0.01
Hospital days per beneficiary (log transformed)										
Baseline year	NA	NA	NA	NA	NA	0.74	0.66			
Months 1-12	NA	NA	NA	NA	NA	0.59	0.34	0.17*** (0.06)	41%	< 0.01
Months 13-24	NA	NA	NA	NA	NA	0.44	0.24	0.12* (0.07)	34%	0.09
Months 1-24	NA	NA	NA	NA	NA	0.79	0.43	0.27*** (0.06)	50%	< 0.01
ED visits, per 1,000 beneficiaries										
Baseline year	1,579	1,381				1,387	1,346			
Months 1-12	1,331	1,045	88 (66)	7.0%	0.18	1,156	1,232	-117 (94)	-9.2%	0.22
Months 13-24	1,180	1,048	-66 (74)	-5.3%	0.37	774	954	-221** (106)	-22%	0.04
Months 1-24	1,255	1,046	11 (62)	< 1%	0.86	1,045	1,151	-146* (87)	-13%	0.09
Outpatient visits, per 1,000 discharges										
Baseline year	8,479	8,614				10,689	11,163			
Months 1-12	7,075	6,629	581*** (188)	8.8%	< 0.01	9,208	8,205	1,476*** (336)	18%	< 0.01
Months 13-24	6,704	5,727	1,112*** (219)	20%	< 0.01	5,490	5,127	836** (425)	15%	0.05
Months 1-24	6,886	6,192	829*** (187)	14%	< 0.01	8,115	7,276	1,312*** (319)	18%	< 0.01
Percentage with any hospitalization										
Baseline year	28	27				36	36			
Months 1-12	20	14	6.1*** (1.2)	43%	< 0.01	31	18	13*** (2.0)	77%	< 0.01
Months 13-24	16	11	4.9*** (1.1)	45%	< 0.01	26	15	11*** (2.6)	70%	< 0.01
Percentage with any ED visit										
Baseline year	63	62				57	57			
Months 1-12	57	50	6.9*** (1.5)	14%	< 0.01	53	53	-0.04 (2.2)	< 1%	0.98
Months 13-24	50	46	4.0** (1.6)	8.7%	0.01	43	50	-6.2** (2.9)	-13%	0.03

Table C.4 (*continued*)

	Site 1					Site 2				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes										
Number of beneficiaries										
Baseline year	1,429	5,279				628	2,637			
Months 1-12	1,428	5,274				628	2,629			
Months 13-24	1,411	5,159				455	1,920			
Months 1-24	1,429	5,279				628	2,637			

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Note: Impact estimates for number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

CRG = Clinical Risk Group; ED = emergency department; NA = not available; NACHRI = National Association of Children's Hospitals and Related Institutions; SE = standard error.

Table C.5. Estimated impact of the NACHRI intervention on select Medicaid use measures during a 24-month follow-up period, by hospital site

	Site 3					Site 4				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries										
Baseline year	926	888				631	574			
Months 1-12	814	595	182 (112)	27%	0.10	543	560	-74 (85)	-12%	0.38
Months 13-24	669	521	111 (116)	23%	0.34	446	385	4.2 (98)	1.0%	0.97
Months 1-24	757	566	155 (104)	26%	0.14	495	479	-41 (83)	-8.0%	0.63
Hospital days per beneficiary (log transformed)										
Baseline year	0.94	0.83				0.60	0.59			
Months 1-12	0.78	0.50	0.18** (0.09)	29%	0.04	0.52	0.54	-0.03 (0.07)	-5.2%	0.69
Months 13-24	0.62	0.41	0.11 (0.09)	23%	0.24	0.39	0.41	-0.03 (0.07)	-6.8%	0.71
Months 1-24	1.0	0.65	0.29*** (0.09)	39%	< 0.01	0.72	0.75	-0.05 (0.07)	-6.2%	0.53
ED visits, per 1,000 beneficiaries										
Baseline year	1,435	1,242				1,698	1,490			
Months 1-12	1,707	1,430	85 (144)	5.2%	0.56	1,476	1,317	-48 (142)	-3.2%	0.73
Months 13-24	1,103	1,170	-260* (148)	-21%	0.08	1,250	1,081	-38 (152)	-2.9%	0.80
Months 1-24	1,477	1,340	-56 (123)	-3.8%	0.65	1,367	1,205	-45 (132)	-3.2%	0.73
Outpatient visits, per 1,000 discharges										
Baseline year	10,181	9,043				7,348	7,696			
Months 1-12	8,812	7,061	613 (446)	7.3%	0.17	7,106	7,103	351 (294)	5.2%	0.23
Months 13-24	5,425	4,308	-21 (530)	< 1%	0.97	5,147	4,917	578* (310)	13%	0.06
Months 1-24	7,524	6,036	349 (424)	4.8%	0.41	6,148	6,059	437 (273)	7.7%	0.11
Percentage with any hospitalization										
Baseline year	45	45				30	30			
Months 1-12	42	26	16**** (3.0)	59%	< 0.01	27	26	0.97 (2.7)	3.7%	0.72
Months 13-24	32	20	12**** (3.2)	61%	< 0.01	22	21	0.69 (2.6)	3.3%	0.79
Percentage with any ED visit										
Baseline year	57	57				65	60			
Months 1-12	60	58	2.1 (2.9)	3.6%	0.48	58	55	3.3 (3.0)	6.0%	0.26
Months 13-24	48	48	-0.19 (3.5)	< 1%	0.96	56	48	7.7** (3.0)	16%	0.01

Table C.5 (*continued*)

	Site 3					Site 4				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes										
Number of beneficiaries										
Baseline year	325	1,356				355	1,299			
Months 1-12	325	1,352				355	1,291			
Months 13-24	267	1,085				352	1,266			
Months 1-24	325	1,356				355	1,299			

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Note: Impact estimates for number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

CRG = Clinical Risk Group; ED = emergency department; NACHRI = National Association of Children's Hospitals and Related Institutions; SE = standard error.

Table C.6. Estimated impact of the NACHRI intervention on select Medicaid use measures during a 24-month follow-up period, by hospital site

	Site 5					Site 6				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries										
Baseline year	865	779				44	38			
Months 1-12	552	407	59 (76)	11%	0.44	36	36	-5.8 (19)	-15%	0.76
Months 13-24	485	346	52 (82)	11%	0.52	50	64	-21 (30)	-35%	0.48
Months 1-24	517	378	52 (73)	10%	0.48	42	49	-13 (19)	-27%	0.49
Hospital days per beneficiary (log transformed)										
Baseline year	0.66	0.59				0.06	0.06			
Months 1-12	0.49	0.33	0.09* (0.05)	21%	0.08	0.05	0.05	0.00 (0.03)	< 1%	0.99
Months 13-24	0.38	0.26	0.04 (0.05)	12%	0.43	0.06	0.06	-0.01 (0.03)	-16%	0.73
Months 1-24	0.72	0.49	0.15*** (0.05)	26%	< 0.01	0.08	0.10	-0.02 (0.03)	-20%	0.49
ED visits, per 1,000 beneficiaries										
Baseline year	1,535	1,389				784	675			
Months 1-12	1,242	1,276	-180* (98)	-13%	0.07	553	567	-122 (75)	-18%	0.10
Months 13-24	1,151	1,193	-189* (104)	-14%	0.07	555	547	-100 (90)	-15%	0.27
Months 1-24	1,199	1,239	-186** (91)	-13%	0.04	549	559	-119 (72)	-17%	0.10
Outpatient visits, per 1,000 discharges										
Baseline year	9,442	8,493				5,720	5,552			
Months 1-12	8,154	7,273	-68 (290)	< 1%	0.81	4,039	4,023	-152 (215)	-3.7%	0.48
Months 13-24	7,108	5,896	263 (332)	3.6%	0.43	4,211	4,135	-92 (271)	-2.1%	0.73
Months 1-24	7,633	6,624	60 (289)	< 1%	0.83	4,118	4,072	-122 (217)	-2.9%	0.57
Percentage with any hospitalization										
Baseline year	36	36				3.2	2.8			
Months 1-12	30	22	8.2*** (2.0)	37%	< 0.01	2.8	3.0	-0.17 (0.98)	-5.8%	0.86
Months 13-24	24	18	5.7*** (1.9)	32%	< 0.01	2.5	3.3	-0.78 (1.1)	-24%	0.46
Percentage with any ED visit										
Baseline year	58	56				43	38			
Months 1-12	53	54	-0.63 (2.2)	-1.2%	0.77	32	35	-3.2 (2.6)	-9.1%	0.22
Months 13-24	49	50	-0.96 (2.2)	-1.9%	0.67	32	31	1.6 (2.8)	5.1%	0.58

Table C.6 (*continued*)

	Site 5					Site 6				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes										
Number of beneficiaries										
Baseline year	677	2,757				422	1,810			
Months 1-12	676	2,748				422	1,798			
Months 13-24	655	2,546				388	1,649			
Months 1-24	677	2,757				422	1,810			

Source: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018.

Note: Impact estimates for number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

*** Significantly different from zero at the .01 level, two-tailed test.

CRG = Clinical Risk Group; ED = emergency department; NACHRI = National Association of Children's Hospitals and Related Institutions; SE = standard error.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the study also estimated the program impacts for the National Association of Children's Hospitals and Related Institutions (NACHRI) using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing the size and the certainty of an impact in a single value. Drawing probabilistic conclusions requires external or prior evidence. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to NACHRI. The prior estimates from HCIA R1 include relatively few child-focused Medicaid awardees, meaning they might not be as comparable to NACHRI as they are to other HCIA Round 2 awardees that focused on adult Medicare populations. The analysis calculated probabilities using the results of a Bayesian regression that jointly models impacts on two core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for two core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for NACHRI led to a Bayesian estimate of the program's impact on hospitalizations of 1 to 2 percent (an estimated increase of three to five hospitalizations per 1,000 beneficiaries per month) and an impact of about 1 percent on emergency department visits (an estimated increase of 3 to 12 visits per 1,000 beneficiaries) across the first two years.

Table D.1. Comparison of frequentist and Bayesian impact estimates for NACHRI in the first two years after enrollment

		Impact estimate (95 percent interval)			Percentage impacts	
Outcome	Follow-up period	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Hospital admissions	Year 1	63 (13, 113)	3.4 (-26, 34)	-2%	16%	< 1%
	Year 2	51 (-4.5, 107)	5.3 (-19, 31)	-1%	16%	2%
ED visits	Year 1	-25 (-102, 52)	3.3 (-95, 105)	-3%	-2%	< 1%
	Year 2	-120 (-206, -34)	12 (-77, 107)	-2%	-10%	1%

Sources: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

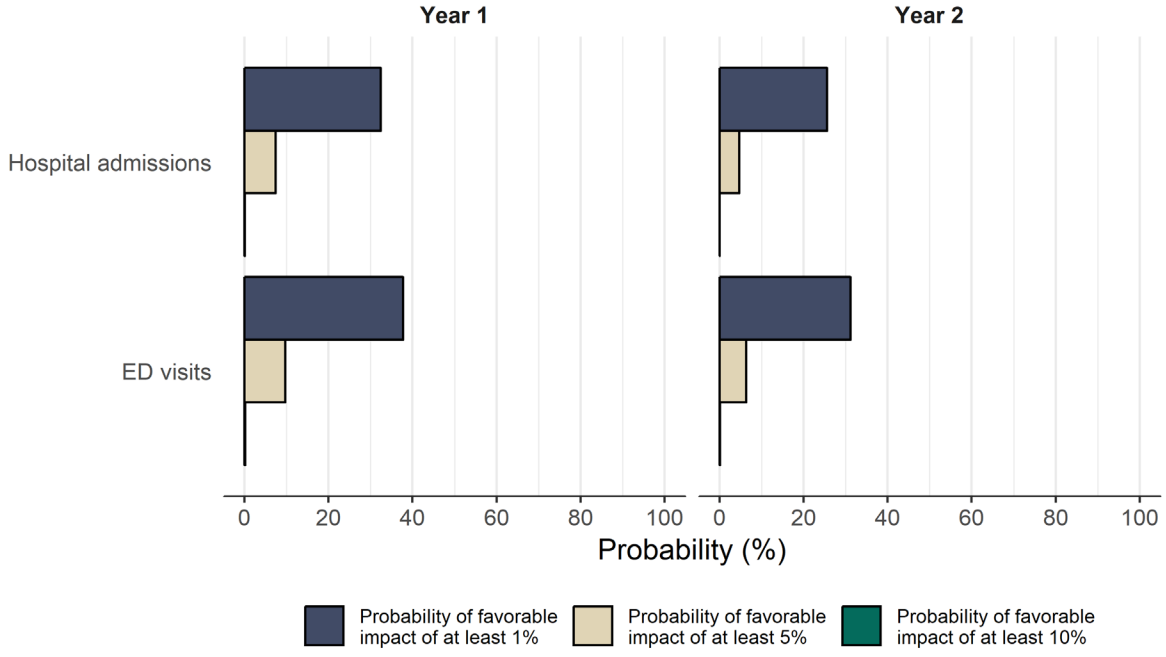
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards; NACHRI = National Association of Children's Hospitals and Related Institutions.

It is unexpected that the direction of the emergency department (ED) visit effect differed between the prior and frequentist analyses and the Bayesian analyses. The Bayesian model estimates the effects on the outcomes combined, rather than separate models for ED visits and hospitalizations. The prior estimates for the ED visit outcome among Medicaid awardees were widely dispersed and had less certainty than the hospitalization outcome. This might have resulted in the Bayesian model putting more weight on the hospitalization outcome in the combined model, which produced more neutral estimates that are more consistent across outcomes. This consistency across outcomes stems from the HCIA R1 data used as a prior, where impacts tend to be consistent across outcomes for the same awardee. To achieve this consistency, the Bayesian model produced estimates that are closer to zero and similar in sign. As a result, the Bayesian impact estimates do not differ meaningfully from zero for any outcome or time period, in contrast to the frequentist Year 1 impact on hospitalizations and Year 2 impact on ED visits, which were significantly unfavorable and favorable, respectively.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in hospitalizations. Figure D.1 shows the probability that NACHRI achieved favorable impacts during each of the first two years on three core outcomes at three different thresholds: a favorable impact of 1 percent or more, 5 percent or more, and 10 percent or more.

Figure D.1. Probability that the NACHRI program had a favorable impact on key outcomes



Sources: Mathematica's analysis of Medicaid claims and enrollment data from April 2013 to November 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Figure D.1 (*continued*)

Note: The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

ED = emergency department; HCIA R1= Round 1 of the Health Care Innovation Awards; NACHRI = National Association of Children's Hospitals and Related Institutions.

There is a modest probability—in the range of 30 to 40 percent—that NACHRI had a favorable impact of 1 percent or more on hospital admissions and ED visits. These probabilities are not large enough to indicate a substantial impact.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: National Health Care for the Homeless Council

September 2020

Brant Morefield, Vladislav Slanchev, Dana Jean-Baptiste, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

NATIONAL HEALTH CARE FOR THE HOMELESS COUNCIL

National Health Care for the Homeless Council (NHCHC) received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to support the Medical Respite Care for People Experiencing Homelessness program. Medical respite care is defined as acute and post-acute medical care provided to homeless individuals who are not sick enough to be in a hospital but who are too sick to recover from a physical illness or injury on the streets. Respite care centers for homeless individuals have proliferated across the nation and the services provided by centers vary widely. In light of these circumstances, NHCHC sought to develop and implement a standard set of respite care services to serve as a model for other programs. The awardee implemented its program in five sites with existing respite care programs for homeless individuals. The target population consisted of people ages 18 and older who experienced homelessness and who were already admitted to a respite care center in one of five Medicaid expansion states (Arizona, Connecticut, Minnesota, Oregon, and Washington State). The program launched in March 2015. Program enrollment ended in May 2017, three months before the end of the intervention period funded by HCIA R2 in August 2017. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The NHCHC sought to implement a standardized enhanced respite care model in existing respite care centers.
- The awardee enrolled adults experiencing homelessness who were admitted to a medical respite care center. The program enrolled 1,441 participants, 1,205 (84 percent) of whom were Medicaid-only beneficiaries and 174 (12 percent) were dually eligible for both Medicare and Medicaid.
- Due to an inability to replicate the eligibility criteria in claims, it was not possible to conduct a rigorous impact evaluation of this program.
- Due to limitations with the state Medicaid data, it was not possible to measure the baseline demographic, health, and service use and spending characteristics of the Medicaid participants.

NHCHC developed and implemented a standard set of respite care services (including care management, patient engagement, and transitional care coordination) that could serve as a model for other programs. The awardee hypothesized that care provided in a medical respite care center that included personalized care management, self-management support, and assistance with transition to primary care would result in better management of chronic conditions, fewer emergency department (ED) visits and hospital stays, and lower health care costs. The goals of the program were to (1) decrease rates of both ED visits and observational stays by 20 percent, (2) reduce the length of stays and cost of the respite care stays by 30 percent (3) decrease 30-day hospital readmissions by 20 percent, and (4) increase use of outpatient services by 10 percent.

It was not possible to conduct a rigorous impact evaluation of the enhanced respite care program because the awardee identified and recruited participants into the program using criteria not available in claims data. Because of the limitations of state Medicaid data, it was not possible to

measure the demographic, health, and service use and spending characteristics of participants before they enrolled in the program. As a result, none of the findings in this report can be used to draw inferences about the impact of the intervention on outcomes.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The purpose of program was to provide a model of comprehensive respite care in safe settings for patients by consistently supplying and tracking the delivery of care management, patient engagement, and transitional care coordination services.
Major innovation	The main innovation of the program was to implement a standardized model of respite care in five sites. The model is defined by the delivery of a core set of services that represent recognized best practices in the field of respite care.
Program components	• Care management involved (1) case management, including providing social services; (2) medication monitoring; and (3) prevention services, including tobacco cessation and influenza vaccination. • Transitional care coordination included providing primary care providers updated health care information and arranging follow-up visits within 7 and 30 days of hospital discharge. • Patient engagement involved motivational interviewing, goal setting, and educating patients.
Target population	The awardee aimed to engage respite care patients ages 18 or older who were homeless. By definition, admission to respite care indicates that the patient had an acute illness or injury. Respite care patients are considered a high-risk and high-cost underserved population. Most of the participants qualified for or were enrolled in Medicaid.
Total enrollment	The awardee enrolled 1,441 patients, representing 46 percent of the original enrollment goal.
Theory of change or theory of action	Access to respite care after a hospitalization for an acute injury or illness would provide participants with a safe environment with access to clinical staff to continue recovery. Furthermore, by providing personalized care management that promoted setting goals and self-management, participants might be better able to manage their chronic conditions, resulting in the use of more preventive services and primary care, fewer ED visits and hospitalizations, and lower health care costs.
Award amount	$2,673,476
Effective launch date	March 2, 2015
Program settings	Medical respite care sites in Arizona, Connecticut, Minnesota, Oregon, and Washington State
Market area	Urban
Target outcomes	• Decrease LOS and cost of respite care stays by 30 percent • Decrease total expenditures • Decrease hospital admissions • Decrease ED visits and observational stays by 20 percent • Decrease hospital readmissions within 30 days by 20 percent • Increase outpatient services by 10 percent • Increase participants' self-management of chronic conditions • Increase smoking cessation efforts • Increase number of participants who receive recommended vaccinations • Increase linkages of participants with social services

Table 1 (*continued*)

Program characteristics	Description
Payment model	At the end of the award, each of the sites had implemented multipayer models that were largely in place before the cooperative agreement. These models varied to reflect differences in Medicaid policies across states and each site's organizational structure and clinical service offerings, including prospective payment system payments for federally qualified health centers; value-based, bundled, or episode payments from managed care or accountable care organization contracts; and/or FFS payments from state Medicaid agencies. All sites had supplemental funding from local hospitals, foundations, and grants.
Sustainability plans	All sites generally maintained core elements of the program, either retaining new staff or incorporating services into existing workflows.

ED = emergency department; FFS = fee-for-service; LOS = length of stay.

PROGRAM DESIGN AND ADAPTATION

The medical respite care program had three key components: (1) care management, (2) transitional care coordination, and (3) patient engagement.[1] Within the framework of a standardized respite care service delivery model, sites had flexibility in how to implement these components. For example, a social worker at one site might have been involved in patient engagement activities, whereas at another site, a nurse could have incorporated goal setting into clinical care. In general, respite staff aimed for daily interactions with participants, but had the flexibility to tailor interactions based on a participant's needs. A staff member might have met with a participant several times a day at the time of admission into respite care, but then tapered visits to two or three times a week by the end of the stay.

Care management

The care management component of the program included delivery of preventive health services, medication management, and case management services by designated staff members. Preventive health services included tobacco cessation counseling and medications to all current smokers, and vaccination of all eligible participants for influenza during the influenza season. Case management staff assessed participants' needs and helped them apply for benefits, including Medicaid, Supplemental Security Income, Temporary Assistance for Needy Families, food stamps, and housing. The staff members who provided care management could vary by respite care location but often a nurse or social worker would meet with participants.

Transitional care coordination

The transitional care coordination component included linking patients to primary care. Program staff obtained hospital records and provided primary care physicians with updated health care information. In respite care sites that shared an electronic medical record with a health system,

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the enhanced respite care stay program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

staff updated the participants' records for follow-up appointments. Staff also arranged follow-up primary care appointments within 7 and 30 days after hospital discharge, and surveyed participants on their experiences with transitioning out of respite care before discharge from respite care. Staff in some respite care sites accompanied participants to their medical appointments while they were in respite care.

Patient engagement

Engaging patients primarily involved motivational interviewing and educating patients to give participants the resources and confidence they needed to set achievable goals for their health and make care plans to meet those goals. Staff promoted engagement through frequent encounters with participants, sometimes increasing the number of clinical staff so that they could see participants at least every other day. The residential setting also enabled regular contact between staff members and participants through one-on-one encounters, support group meetings, and educational sessions about tobacco cessation and access to benefits. The presence of these services in respite care centers helped to engage patients in other program components, such as preventive health services and follow-up care on mental health issues.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee was somewhat successful in providing services as intended. Its partial success was due in part to high staff motivation and low staff turnover. Staff dedicated to working with a homeless population enhanced the awardee's ability to effectively deliver services, resulting in low staff turnover during the cooperative agreement. The staff at each site was collaborative, sharing workloads and frequently checking in through team meetings about the intervention. Trainings on motivational interviewing and self-management gave staff the tools to engage participants, and established relationships between the respite care center and the local primary care practices facilitated coordinating transitional care. Overall, program leaders and staff felt that the program made a difference in meeting critical needs of this vulnerable population and positively affected participants' health goals. In a survey of respite care staff, all respondents reported that the program had a positive impact on participants' satisfaction and quality of life.

However, several implementation challenges limited the program's overall effectiveness. First, NHCHC had difficulty meeting its enrollment goal due to fewer eligible participants than expected. Participants were enrolled in the program only after they were admitted to one of the participating respite care sites. However, respite care patients remained in respite care longer than expected, which led to a shortage of available beds and reduced the total number of potential participants. The awardee expected respite care stays to last two to four weeks, but the average length of stay during the cooperative agreement across sites was seven weeks. The awardee tried to address this challenge in several ways. One site expanded the number of respite beds by partnering with a local shelter. Another site improved care transition processes to reduce

the length of respite stay. A few sites attempted to reduce the number of inappropriate referrals of patients who were ineligible for respite care because they required a higher level of care.

Second, staff faced challenges enrolling respite care patients into the program and obtaining their consent to share data. After the program admitted a patient to respite care, a staff member would approach him or her about consenting to share data about services received, completing surveys for evaluation purposes, and obtaining claims data. Potential participants had concerns about confidentiality and wariness over joining a research study or government-funded program. Program staff attempted to overcome mistrust using several strategies, such as using clinical or master's level-trained staff to help at the hospital referral stage, discussing successful enrollment strategies during team meetings, building rapport with patients over a few days before attempting to enroll them, and explicitly describing to patients how the program would use their data.

Implications of program implementation for achieving program goals

- The awardee was partially successful in providing services as intended, due in part to high staff motivation and low staff turnover, and achieved high rates of physician follow-up.
- Despite dedicated, committed program staff, the difficulties with enrolling and engaging patients underscored the challenges of working with this vulnerable population.

A third implementation challenge included a lack of specific types of staff to meet service delivery goals, such as independent medical providers to offer first-line pharmacotherapy for smoking cessation or influenza vaccination, or a program manager with broader authority over implementation, rather than the narrowly defined data manager role. Some sites lacked a dedicated case manager, such that one staff member might have had multiple duties beyond traditional case work (for example, substance abuse counseling).

Fourth, there were substantial amounts of missing data for implementation measures, particularly for patient-centered measures of engagement. Program staff suggested that uniform training of all respite care staff might have helped mitigate this challenge. Finally, the medical and social complexity of the participants' circumstances prevented some participants from engaging fully in the program. Participant could clinically deteriorate quickly, requiring a return to the hospital or participants with substance abuse could leave against medical advice before engaging in respite care services.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Enrolling and engaging participants

The awardee had originally planned to enroll 3,127 participants, but achieved only 46 percent (1,441) of this target by May 2017 when it stopped enrolling new patients. Of these, 84 percent

(1,205) were covered under Medicaid and 12 percent (174) were dually eligible for both Medicare and Medicaid. Only 3 percent (42) were Medicare-only beneficiaries and 1 percent (20) was uninsured or had unknown insurance status. The program expected a large proportion of Medicaid participants due to all sites being located in Medicaid expansion states.

The awardee was most successful providing transitional care coordination services. Sixty-five percent of participants had a primary care follow-up visit within 7 days of enrollment and 96 percent had a visit within 30 days. In terms of the care management component, one-third of participants received housing assistance during their respite care stay, and two-thirds of those who smoked received a pharmacotherapy intervention such as nicotine replacement therapy or bupropion. Some case management services were needed less or were more difficult to implement than anticipated. For example, many participants had already enrolled in Medicaid (93 percent) and received food stamps (68 percent) and, therefore, did not need assistance accessing these programs. Similarly, nearly half had already received an influenza vaccination before entering respite care and one-third had received a flu shot. Nearly 40 percent of participants did not complete the surveys on patient self-management, health literacy, and care transition experience, underscoring the difficulty in engaging this patient population.

Due to problems with the quality and completeness of the Medicaid data files for the five states that implemented the program, it was not possible to measure the baseline demographic, health, service use, and spending characteristics of the Medicaid participants. Although the awardee enrolled a small number of Medicare beneficiaries, it primarily intended the intervention to serve a Medicaid population; this analysis does not present the baseline characteristics of the small subgroup of Medicare beneficiaries.

Challenges of estimating program impacts

A rigorous impact analysis was not conducted for this program due to (1) the inability to reliably identify homeless beneficiaries in Medicaid claims data, (2) the use of clinical judgment to refer patients to respite care from community settings and hospitals, and (3) concerns about the quality and completeness of state Medicaid data. First, staff reviewed awardee-supplied state Medicaid data for participants for encounter codes indicating homelessness. The proportion of participants with a homeless encounter code in their Medicaid files varied widely, ranging from 28 percent at the respite care site in Arizona to 81 percent at the Connecticut site. Second, the clinical judgement of the referring provider determined eligibility for respite care and there was not a standard set of diagnoses used to guide referrals across the sites. Furthermore, program staff waited for a few days to enroll respite care patients into the intervention to avoid enrolling patients who would leave respite care against medical advice, which might have led to a selection bias of participants more likely to engage and benefit from program services. Finally, additional limitations of state Medicaid data files received from the awardee included the absence of an eligibility file for two states; the lack of state Medicaid data from two states, and limited follow-up data for the remaining three states. As a result, these data could not be used to estimate program impacts.

CONCLUSION

The NHCHC was partially successful in implementing its standardized respite care model in sites across five Medicaid expansion states. The program hired staff dedicated to working with the homeless, achieved high rates of primary care follow-up, and surpassed national rates of smoking cessation therapy. Less successful elements of implementation included lower than expected enrollment, lower need for certain services such as case management for obtaining Medicaid (which might be specific to centers in Medicaid expansion states), and difficulty collecting self-reported data from this vulnerable patient population. The inability to identify treatment beneficiaries and replicate the eligibility criteria for a comparison group in administrative data prevented a rigorous assessment of the program's impacts on outcomes.

PROGRAM SUSTAINABILITY

Although the NHCHC ended oversight and monitoring of its medical respite services program at the end of its award in August 2017, all five participating sites continued to deliver program services with support from a variety of payment models. These sites maintained the core elements of the program by keeping new staff or embedding new services into existing staff workflows and electronic health records, if available. Many sites retained their program data managers for quality improvement initiatives. They modified other parts of the program (for example, the participant surveys) to reduce costs and work better for their particular contexts and participants. One site added another local respite care program funded by local hospitals.

All five participating sites continued using multipayer payment models that were largely in place before the cooperative agreement. Sites affiliated with federally qualified health centers (FQHCs) received an enhanced prospective rate from the state Medicaid agency. They also received additional payments for respite care services through multiple Medicaid-contracted managed care organization (MCO) agreements for Medicaid managed care enrollees, and via a flat monthly payment per beneficiary for respite care services provided by Medicaid- or Medicare-contracted accountable care organizations. Sites not affiliated with an FQHC had either multiple Medicaid-contracted MCO agreements, receiving payments for respite care services through

NHCHC's proposed payment model

- Each of the five sites that participated in the NHCHC's program adopted its own payment model. These models varied to reflect differences in Medicaid policies across states, as well as differences in each site's organizational structure and clinical service offerings.
- The sites relied on multipayer models that pre-dated the cooperative agreement. Those affiliated with FQHCs had enhanced prospective payment system payments and MCO payments that included respite services. Non-FQHC sites received fee-for-service rates from either an MCO or, in one state's case, under a subcontract from the patient's medical provider. One site received payments from Medicare and Medicaid accountable care organizations.

a daily rate arrangement or received FFS payments for respite care services from the patient's medical provider from the Medicaid state agency.

All sites supplemented the payment model through additional funding from grants, foundations, and partnering hospitals that stood to save money by transferring patients from inpatient to subacute care settings. Sites reported that some hospitals decreased their financial support for the program as they observed an increase in Medicaid funding support, so sites considered applying for grants from hospital community benefit programs. The awardee attempted to help sites in negotiations for additional payment arrangements (for example, shared savings) by estimating the financial value of the program. However, this turned out to be difficult because the cost and method of providing services varied at each site, and nondisclosure agreements with some MCOs prevented the disclosure of itemized costs per service for value-based or bundled payment codes. As a result of these constraints, none of the participating sites had been able to execute additional payment arrangements by the end of the award.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Nebraska Medicine

September 2020

Jia Pu, Colene Byrne, Alison Banger, Bob Schmitz, and Randall Brown

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

NEBRASKA MEDICINE

Nebraska Medicine, an academic medical center and teaching hospital for the University of Nebraska Medical Center (NMC), received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to develop and test the Remote Interventions Improving Specialty Complex Care (RIISCC) program. RIISCC provided remote patient monitoring (RPM) for participants with diabetes for 90 days after their discharge from the hospital. The target population consisted of patients with a diagnosis of type 2 diabetes who had an inpatient admission or outpatient visit at Nebraska Medicine or Nebraska Medicine-Bellevue Hospital and, starting in Year 3, from other local hospitals. The program launched in December 2014, four months after award. The intervention period covered under HCIA R2 ended in August 2017. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The RIISCC program aimed to help patients with type 2 diabetes improve their self-management and reduce unplanned readmissions and ED visits by providing remote patient monitoring and health coaching. A small financial incentive encouraged patients to participate.
- A critical component of the program was daily in-home telemonitoring of blood glucose, body weight, and blood pressure.
- A registered nurse provided coaching and education via telephone each week for three months and every month for an additional nine months.
- The impact analysis was based on 430 Medicare fee-for-service (FFS) beneficiaries who enrolled in the program (23 percent of total enrollees) and a matched comparison group of 1,855 beneficiaries who received care from other large nonparticipating hospitals in Omaha.

The program's underlying assumption was that health coaching—using daily information on enrollees' weight, blood pressure, and blood glucose levels—could improve patients' management of diabetes, their care and health behaviors. Participants submitted daily readings and worked closely with their health coaches to set and work toward goals related to diet, exercise, medication adherence, and weight and blood pressure control. The health coaches provided guidance and educated patients at a pace and in an environment that was conducive to learning. For example, the coaches often learned from RPM data that patients were not compliant with taking insulin and other medications, sometimes due to medication costs. They often directed patients to Nebraska Medicine's outpatient pharmacy, which has several programs that enable participants to apply for assistance to receive free or reduced-cost insulin. The health coaches also identified and followed up on potential changes in patients that might indicate a need for participants to see their physicians soon or that could indicate a potential emergency. The awardee expected these improvements, in turn, to result in lower unplanned inpatient and ED service use, increased regular contact with primary care physicians, reduced total costs of care, and higher quality of life. The program was available to adults ages 19 and older who resided in Douglas, Sarpy, or Cass counties in Nebraska and met the enrollment criteria.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Nebraska Medicine provided RPM and health coaching to individuals with type 2 diabetes who had an inpatient admission or outpatient visit to Nebraska Medicine and Nebraska Medicine-Bellevue Hospital and who resided in targeted areas of the greater Omaha, Nebraska, metropolitan area.
Major innovation	The program used RPM and health coaching to improve patients' self-management of diabetes.
Program components	• Patient engagement and coaching • Health IT and telemedicine
Target population	The program sought to engage residents of Douglas, Sarpy, and Cass counties, Nebraska, diagnosed with type 2 diabetes and who had an inpatient admission or outpatient visit, for any reason, at Nebraska Medicine or Nebraska Medicine-Bellevue Hospital.
Participating providers	Nebraska Medicine, Nebraska Medicine-Bellevue Hospital, Bellevue Clinic, Charles Drew Community Health Center
Total enrollment	The awardee enrolled 1,903 patients in RIISCC from September 2014 through August 2017, representing 76 percent of its original enrollment goal.
Level of engagement	Among the surveyed staff, 41 percent strongly agreed and 24 percent somewhat agreed that the program had successfully engaged participants. However, Nebraska Medicine's internal analyses indicated that the program might not have been as successful in engaging its higher-risk patients.
Theory of change/ theory of action	By providing early and timely post-discharge health coaching and RPM and incentives in the form of gift cards for patients to participate (a $10 gift card at each stage of the program and an additional $10 gift card for returning their equipment, potentially $50 in all), patients will better self-manage their diabetes and keep their conditions stable, which will result in reduced inpatient and ED use and associated costs due to uncontrolled diabetes.
Award amount	$9,993,626
Effective launch date	December 22, 2014
Program settings	Academic medical center, community health care clinics, and participants' homes
Market area	Douglas, Sarpy, and Cass counties in Nebraska
Target outcomes	• Improved blood pressure and hemoglobin A1c • Increased regular contact with primary care physicians • Reduced total costs of care for participants • Fewer all-cause unplanned hospital admissions and readmissions • Fewer ED visits • Reduction in body mass index
Payment model	New FFS payment with shared savings and bundled per-episode payment for telehealth services
Sustainability plans	The awardee focused on engaging organizational leaders as its sustainability strategy in the third program year, including outlining the program's benefits relative to its cost and planning to meet with organizational leaders and payers to discuss funding and sustaining the program.

ED = emergency department; FFS = fee-for-service; IT = information technology; RIISCC = Remote Interventions Improving Specialty Complex Care; RPM = remote patient monitoring.

The impact analysis was limited to 430 Medicare FFS beneficiaries who had an inpatient admission or outpatient visit at Nebraska Medicine and Nebraska Medicine-Bellevue hospital,

had a diagnosis of type 2 diabetes and met the other study inclusion criteria (that is, age 19 or older, not pregnant, English speaking, able to understand and use the equipment, and residing in the targeted geographic area). These beneficiaries represented fewer than 10 percent of Medicare beneficiaries who were potentially eligible for the program based on a review of claims data and met the claims-based inclusion criteria for the study.[1] The comparison group included 1,855 Medicare FFS beneficiaries with similar demographic and health characteristics who had inpatient or outpatient care at nonparticipating hospitals in Omaha during the intervention period and thus were ineligible to participate in the program. Table 2 summarizes the key features of the evaluation. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Features of program evaluation

Features	Description
Evaluation design	The impact estimates relied on a difference-in-differences study design. This design measures program effects as the (regression-adjusted) change in outcomes among the intervention group before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. To control for the large differences in mortality between the two groups, the regression analysis included a control variable for whether the beneficiary died during the follow-up period.
Intervention group for evaluation	The intervention group for the evaluation relied on 430 Medicare FFS beneficiaries (among the total enrollment of 1,903) who enrolled in the RIISCC from September 2014 to August 2017, representing 23 percent of total enrollment during that period. The study intervention group does not include an estimated 1,346 patients who either were not enrolled in Medicare or could not be identified in the Medicare enrollment database. It also excluded 94 beneficiaries who did not meet the claims-based eligibility criteria for the study, such as being enrolled in Parts A and B. In addition, the evaluation dropped 27 beneficiaries without a valid anchor encounter, 4 beneficiaries who died within 30 days of enrollment, and 2 outlier beneficiaries.
Comparison group	The comparison group included 1,855 Medicare FFS beneficiaries with similar demographic and health characteristics who received inpatient or outpatient care at any other nonparticipating hospital in Omaha during the intervention period and thus were ineligible to participate in the program.
Limitations	If participants differed from eligible nonparticipants in meaningful ways not captured in Medicare administrative data, the impact estimates could be biased. The low participation rate (less than 10 percent) would make it virtually impossible to identify impacts if measured over all eligible beneficiaries.

FFS = fee-for-service; RIISCC = Remote Interventions Improving Specialty Complex Care.

PROGRAM DESIGN AND ADAPTATION

The RIISCC program service delivery model had two main components: (1) patient engagement and coaching, and (2) health information technology (health IT) and telemedicine.[2]

[1] The claims-based inclusion criteria are a subset of all enrollment criteria used over the course of the project. Other inclusion criteria, such as English speaking and ability to understand and use the equipment, cannot be replicated by using Medicare FFS enrollment and claims data. Nebraska Medicine reported that almost half of those they approached agreed to participate.

[2] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Engaging and coaching participants

Participants received weekly calls from health coaches for three months to discuss critical test values and arrange for tests or consultations as needed. After 90 days of RPM, participants went to one of the primary care clinics, returned their RPM equipment, had their hemoglobin A1c checked, received a foot and eye exam, and received nutritional counseling. After the 90 days, participants received monthly calls from a health coach for an additional nine months.

The awardee considered engaging patients as the critical component for the program's success and a driver for helping patients improve outcomes, such as improved A1c levels and medication adherence. Program staff described a wide array of approaches to working with participants aimed at meeting the participant where he or she was. For example, a clinical diabetes educator noted that a goal to walk 10,000 steps a day might be feasible for a younger person with a regular workday job and without significant family responsibilities, but might simply discourage someone who worked two jobs and had caregiving responsibilities at home. The health coaches had consistent contact with participants and took the time to help them understand the challenges they faced managing their diabetes, as well as to identify behavioral changes and approaches to addressing those challenges.

Health IT and telemedicine

A medical assistant installed RPM equipment in each participant's home and demonstrated its use to the participant. The equipment transmitted information about the participant's weight, blood pressure, and blood glucose values to the health coach and providers. In Year 2, participants also had the option to go to Nebraska Medicine's telemedicine hub to learn how to use the equipment before taking it home. To support the flow of information, the awardee developed an interface enabling patients' data from the RPM system to transfer into the electronic medical record (EMR) system at Nebraska Medicine and at Nebraska Medicine-Bellevue Hospital. The interface facilitated communication and collaboration between the health coaches and participants' primary care providers (PCPs). The EMR system set aside a new section for telemedicine encounters; it displayed the RPM data along with the coaching notes. Clinicians involved in the program were highly satisfied with the interface, which supported their clinical workflows. Nebraska Medicine's charting improved significantly over the three years. Nebraska Medicine implemented more structure and added smart phrases to its EMR, which enabled the health coaches to enter data more quickly and serve more patients. Nebraska Medicine also compiled and electronically linked not only the data available from the EMR and RPM systems but the various types of program data that staff collected as well. This integration across its information systems ensured a more comprehensive view of a participant's progress and care.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

After initial difficulties related to unexpected staff departures, the awardee began to deliver intervention services as intended, including installing RPM equipment and providing weekly coaching to participants. The awardee provided consistent services to participants, engaged local providers, achieved high levels of participant satisfaction, and, according to its own clinical performance metrics, helped participants control their diabetes. The program provided an additional level of support for participants, enabling physicians to focus on direct patient care.

Implications of program implementation for detecting impacts

- Nebraska Medicine reported difficulties with enrollment, disenrollment, and engaging patients, particularly for the sickest and highest-risk patients. Because the highest-risk patients were less likely to enroll in the program and more likely to disenroll, meeting some of the targeted outcomes may have been challenging, as those enrollees would likely benefit more than others from improving their management of diabetes. Improved self-management might have helped them reduce their high rates of hospitalizations and ED visits.
- The sample of Medicare beneficiaries enrolled in the program and used in this analysis represented less than 10 percent of the beneficiaries who appeared to be eligible to participate and met the claims-based inclusion criteria. Therefore, a primary concern of the study is that the experience of beneficiaries who enrolled in the program might not generalize to the full target population.
- The small number of Medicare enrollees means that the study was likely to detect only large program effects.

However, the awardee faced significant challenges recruiting patients into the program. Nebraska Medicine initially planned to recruit patients before their discharge from the hospital, but this proved infeasible due to a number of factors. For example, many patients were discharged before program staff could meet with them and explain the program. Also, patients were often overwhelmed with information at discharge and not ready to decide about participating in the program. It therefore became necessary to recruit participants by telephone after they had returned home from their hospital admission or outpatient visit, which made recruitment much harder to achieve. In November 2016 (more than two years after award), the awardee reduced its target enrollment from 3,300 to 2,300. Nebraska Medicine almost reached that goal by the end of the program in August 2017—total enrollment over the course of the program was 1,903. The program also experienced problems with disenrollment. About one-quarter of enrollees disenrolled before they completed the 90-day program. Nebraska Medicine analyzed its enrollment and disenrollment data to better understand potential factors that might have influenced patients to enroll and remain in the program. Most patients who disenrolled did so within four weeks of enrolling. Some of the reasons reported for withdrawing from the program include loss of glucose meters, moving to another state, accidents, too busy to participate, and not receiving requested information.

ESTIMATING PROGRAM IMPACTS

Study sample

The study relied on 430 Medicare FFS beneficiaries who enrolled in the RIISCC program from September 1, 2014, through August 31, 2017 (representing 23 percent of the total 1,903 patients who enrolled during that period), and 1,855 matched comparison beneficiaries treated only in other Omaha hospitals (and thus were unable to participate in the program) during the same period. The study intervention group did not include an estimated 1,346 patients who either were not enrolled in Medicare or could not be identified in the Medicare enrollment database. The study sample also excluded 94 beneficiaries who were enrolled in Medicare Advantage, were not enrolled in both Parts A and B of Medicare, or for whom Medicare was not the primary payer. Another 33 beneficiaries were excluded based on other study exclusion criteria: for example, died within 30 days of enrollment, did not have a valid anchor encounter, or classified as an outlier based on participant's baseline characteristics (Appendix A, Table A.1).

Recruiting, enrolling and engaging participants

The Nebraska Medicine and Nebraska Medicine-Bellevue Hospital's EMR system, Epic, facilitated implementing the program and enrolling participants. Each day, a computer program based on a set of program eligibility criteria generated a list of eligible patients who had an inpatient admission or outpatient visit at Nebraska Medicine and Bellevue. After the program identified patients as eligible, they had an in-person meeting with a lead nurse in the hospital (if possible), who briefly introduced the program and gave the patient information about the program to take home. After returning home, the program mailed the same information to the patient's home, and health coaches, who were registered nurses, followed up by telephone to try to recruit patients into the program. In addition, primary care medical teams and diabetes educators in both the inpatient and outpatient areas encouraged program participation during patients' hospitalizations and initial follow-up visits after discharge. Patients also received invitations to participate via messages sent through Nebraska Medicine's patient portal. Nebraska Medicine estimated that about half of those they approached agreed to participate, regardless of the insurance type. If the patient agreed to participate, a medical assistant arranged to go to the participant's home to deliver, install, and demonstrate use of the RPM equipment. In the last year of the program, participants also had a second option and could visit Nebraska Medicine's telemedicine hub to learn how to use the equipment before taking it home.

According to program staff and leaders, Nebraska Medicine partly succeeded in engaging program participants. Among the surveyed staff, 41 percent strongly agreed and 24 percent somewhat agreed that the program successfully engaged participants. Almost three-quarters (71 percent) felt that engaging participants in the program was one of the most helpful factors for achieving program goals. However, Nebraska Medicine's internal analyses indicated that the program might not have been as successful engaging higher-risk patients (for example, those with higher blood pressure or A1c levels and smokers). Higher-risk patients were more likely to

disenroll than lower-risk patients. For example, Nebraska Medicine found that patients with diabetes and poor control (blood glucose A1c greater than 9) were 58 percent more likely to disenroll than patients with diabetes and good control (blood glucose A1c 9 or lower), smokers were 70 percent more likely to disenroll than nonsmokers, and patients with high blood pressure were 35 percent more likely to disenroll than patients who did not have high blood pressure. Some patients reported they had already participated in other diabetes management programs or were overwhelmed by the disease and thus could not participate. It is also possible that these patients might have had other more pressing health issues and thus could not focus on diabetes management.

Characteristics of treatment and comparison group and eligible nonenrolled beneficiaries

Overall, the RIISCC program enrolled a group of beneficiaries with high needs. The treatment group beneficiaries by definition had at least one chronic medical condition—that is, diabetes. Moreover, all beneficiaries had at least one inpatient admission or outpatient visit to Nebraska Medical or Nebraska Medicine-Bellevue hospital during the baseline year. It is therefore not surprising that their Medicare service use and expenditures were well above average. Mean total Medicare expenditures per beneficiary per month for the 430 beneficiaries who enrolled in the program was $1,586 during the baseline year—almost double the national average of $792. Acute hospital admissions, more than 800 per 1,000 beneficiaries per year, were almost three times the national average of 274 per 1,000 beneficiaries in 2014. ED visits and observation stays were also well above national averages (Table 3).[3]

The 430 enrolled beneficiaries represent less than 10 percent of Medicare beneficiaries who were potentially eligible for the program and met the claims-based inclusion criteria. The claims-based inclusion criteria are a subset of all of the enrollment criteria Nebraska Medicine used over the course of the project.[4] To assess differences between the enrolled and eligible nonparticipants, the evaluation compared baseline characteristics between the enrolled beneficiaries (N = 430) and beneficiaries who received inpatient or outpatient care from the program hospitals with a diagnosis of type 2 diabetes but not enrolled (N = 5,084). The evaluation defined the baseline year as the 365 days before each beneficiary's enrollment date. It defined the enrollment date as the date of a participant's first face-to-face encounter with a health coach. A pseudo-enrollment date was assigned to each comparison beneficiary and eligible nonparticipant based on the date of the anchor encounter, which is an inpatient discharge or outpatient visit at one of the program hospitals (for eligible nonparticipants) or other Omaha hospitals (for comparison beneficiaries).

[3] For national average rates, see the Centers for Medicare & Medicaid Services' "Public Use File; New Data on Geographic Variation," available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.

[4] The evaluation estimate understates the participation rate among those that the program viewed as meeting all of their eligibility criteria, for two reasons. First, this analysis could not impose some of the criteria used by NMC, such as whether English-speaking and ability to understand and use the equipment. Second, one of the participating hospitals was not a referral source until late in the three-year program period.

Compared to the eligible nonparticipants, enrolled beneficiaries were more likely to be eligible for Medicare based on disability and therefore more likely to be younger than 65 years (Table 3). They were also more likely to be dually enrolled in Medicare and Medicaid. In addition, hospitalization and ED use in the 12 months before enrollment were substantially higher for enrolled than for eligible non-enrolled beneficiaries. On the basis of pre-enrollment information, there is little or no evidence of favorable selection into the sample. In fact, the baseline data suggest that participants were sicker on average than eligible nonparticipants (indicating adverse selection into the program). However, mortality during the 12 months after enrollment among eligible nonparticipants was much higher at 8.5 percent than the 3.7 percent rate observed for participants. These findings indicate that from the set of eligibles, the program tended to enroll beneficiaries who on average were higher than average past users of inpatient and ED use, but who were much less likely to be terminally ill.

A primary concern associated with the observed differences is that enrolled beneficiaries might have differed from eligible nonparticipants in unmeasurable ways, and thus selecting a comparison group that matched well on pre-enrollment characteristics and service use captured in Medicare administrative files and claims might not adequately control for the potential selection bias. This concern was validated by comparing treatment and matched comparison groups on characteristics at baseline and mortality rates. Although the two groups were well balanced on all baseline variables, including expenditures, subsequent mortality among comparison group beneficiaries was much higher than treatment group beneficiaries (6.4 versus 3.7 percent) (Table 3). Because it is implausible that the program would have any effects on mortality in a single year, and certainly not such a large effect, the comparison suggests that enrollees as a group were probably at lower risk of adverse outcomes stemming from unmeasured differences, such as having a terminal illness, than the comparison group.

Table 3. Baseline characteristics of treatment and comparison group Medicare beneficiaries

Characteristic	Treatment group (N = 430)	Comparison group (N = 1,855)	Eligible nonparticipants (N = 5,084)
Demographics			
Age at enrollment, years	68	69	69
Age group, %			
Younger than 55	10	9	10
55 to 64	18	17	11
65 to 69	29	25	26
70 to 79	34	36	36
80 and older	9	13	17
Male, %	44	43	48
White, %	78	78	79
Original reason for Medicare eligibility, %			
Original reason for Medicare entitlement, disability[a]	43	40	28

Table 3 (*continued*)

Characteristic	Treatment group (N = 430)	Comparison group (N = 1,855)	Eligible nonparticipants (N = 5,084)
Dually eligible for Medicare and Medicaid	25	25	22
Health status and diagnosis, %			
COPD	26	25	19
CHF	25	28	22
Diabetes without complication	34	34	45
Diabetes with chronic or acute complications	66	66	55
Hypertension	94	95	90
Morbid obesity	16	17	13
Vascular disease	20	20	19
HCC score[b]			
Mean	2.0	2.0	1.9
25th percentile	1.1	1.0	0.8
Median	1.7	1.6	1.3
75th percentile	2.6	2.6	2.4
Service use and expenditures during the year before enrollment			
Any hospitalizations, %	50	50	41
Any outpatient ED visits, %	62	52	36
Number of hospital admissions (per 1,000 beneficiaries)	835	888	739
Number of outpatient ED visits (per 1,000 beneficiaries)	1,448	1,129	796
Total Medicare expenditures ($ PBPM)	1,586	1,595	1,642
Mortality within 12 months of enrollment, %	3.7	6.4	8.5

Sources: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2017.

Notes: The evaluation defined the baseline year as the 365 days before each beneficiary's enrollment date. It defined the enrollment date as the date of a participant's first face-to-face encounter with a health coach.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

A pseudo-enrollment date was assigned to each comparison beneficiary and eligible nonparticipant based on the date of the anchor encounter, which is an inpatient discharge or outpatient visit at one of the program hospitals (for eligible nonparticipants) or other Omaha hospitals (for comparison beneficiaries).

None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level, 2-tailed test.

Appendix B presents the full balance results. Exact matching variables include age category, sex, anchor encounter type, and dual Medicare–Medicaid eligibility.

[a] Includes participants with both a disability and ESRD.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disease; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact estimates rely on a difference-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables will capture differences on unobserved variables as well. The primary outcomes were total Medicare spending and spending on hospitalizations, number of hospital admissions and primary care visits, percentage of beneficiaries with an ED visit, and percentage with a hospital readmission within 30 days after a discharge.

The large observed differences in mortality between enrollees and non-enrolled eligibles raised major concerns about unmeasured selection bias (Table 3). Beneficiaries who die in a given year have average costs that are six times higher than the average for all Medicare beneficiaries. Thus, the evaluation estimated program impacts controlling for mortality within 12 months of enrollment.

The evaluation defined the pre-enrollment period as the year before each participant's enrollment date and the post-enrollment period as the one year after enrollment. It defined the enrollment date as the date of a participant's first face-to-face encounter with a health coach as reported by the awardee. As previously noted, a pseudo-enrollment date for defining the pre- and post-enrollment periods was assigned to each comparison beneficiary and eligible nonparticipant based on the discharge date of the anchor encounter. Appendix A describes the statistical models used to estimate the effects of the program, and it identifies the final analytic sample.

IMPACT RESULTS

Table 4 shows estimated program impacts with and without mortality as a control variable. Before controlling for mortality, results indicate the RIISCC program had a favorable impact on total Medicare expenditures. Total expenditures increased more for the comparison beneficiaries than for the enrolled beneficiaries between the baseline and one-year follow-up period, leading to a statistically significant estimated *reduction* of **16 percent** (Table 4). An estimated **20 percent** decrease in inpatient expenditures mainly drove the relative decrease in the total expenditures in the program year. The number of hospitalizations during the year after enrollment, as well as the proportion of beneficiaries who had a 30-day readmission, also fell more among participants relative to comparison beneficiaries.

However, after controlling for the sizable mortality difference, the results suggest that these estimates substantially overstate true program effects. The estimated program impact on total Medicare expenditures dropped by about 40 percent after controlling for mortality, from 16 to 11 percent, validating the concern that favorable selection into the program led to biased (favorable)

estimates of program effects. Similarly, the estimated effect on number of hospitalizations dropped from 9 percent of the mean to essentially zero after controlling for mortality. The estimates suggest that the program had no discernible favorable effects on hospitalizations or other service use, further validating the concern that the sizable (but not statistically significant) estimated cost difference between treatment and comparison groups does not represent a true program effect.

Adding mortality as a control variable did not change the results for ED visits and primary care visits in ambulatory settings. The findings suggest that the RIISCC program had an adverse impact on the likelihood of having an ED visits, but had no effect on the number of ED visits (see Appendix C). In the one-year follow-up period, the difference between the treatment and comparison group beneficiaries indicates that the proportion of beneficiaries who had an all-cause ED visit *increased* by an estimated 7 percentage points or 17 percent for the treatment group. The program was also associated with an 8 percent increase in primary care visits in ambulatory settings. Both estimated increases were statistically significant. The discussion section explains how such impacts might have arisen. Appendix C presents the full results of the impact analysis. Appendix D provides the results from a Bayesian analysis.

Controlling for mortality accounts for treatment-comparison differences in the proportion of sample members that had a terminal illness, but does not account for other possible differences between enrollees and the comparison group on factors that affect future costs but are not reflected in claims data. For example, beneficiaries whose health was deteriorating rapidly but did not die might also have been less likely to enroll than eligible beneficiaries whose health problems were more transitory. The underlying differences in the cost trajectories of participants and comparison beneficiaries due to unmeasured pre-existing differences between the two groups cannot be adequately captured by differences in pre-enrollment service use or other data observable from claims. Thus, the estimated effect on expenditures may potentially overestimate the true program impact, even when controlling for mortality.

Table 4. Estimated impact of RIISCC on selected outcomes, in 12-month follow-up period

	Without controlling for mortality	Controlling for mortality
Expenditures PBPM[a]		
Impact ($)	-$291	-$186
Percentage impact	-16%	-11%
p-value	0.02	0.12
Acute inpatient payments[a]		
Impact (rate)	-$150	-$80
Percentage impact	-20%	-12%
p-value	0.06	0.30
Number of hospitalizations, per 1,000 beneficiaries[a]		
Impact (rate)	-70	-9.9

Table 4 (*continued*)

	Without controlling for mortality	Controlling for mortality
Percentage impact	-9%	-1%
p-value	0.33	0.89
Number of primary care visits in ambulatory settings, per 1,000 beneficiaries[a]		
Impact (rate)	535	545
Percentage impact	8%	8%
p-value	0.05	0.05
Percentage of beneficiaries with a 30-day readmission		
Impact (rate)	-2.9	-2.2
Percentage impact	-22%	-17%
p-value	0.09	0.20
Percentage of beneficiaries with an outpatient ED visit		
Impact (rate)	6.9	7.4
Percentage impact	16%	17%
p-value	0.01	0.01
Sample size		
Treatment	430	430
Comparison	1,855	1,855

Source: Mathematica's analysis of information from Medicare claims and enrollment data from September 2014 through August 2017, as of March 13, 2019.

Notes: Impact estimates rely on the regression-adjusted difference between the randomized treatment and control group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the predicted treatment group mean in the post period minus the impact estimate). Appendix C presents the full impact estimates. Appendix D shows the results from the Bayesian analysis.

[a] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

ED = emergency department; PBPM = per beneficiary per month; RIISCC = Remote Interventions Improving Specialty Complex Care.

Overall, the study results suggest that the program did not have discernable effects on the core outcomes, including the total Medicare expenditures, number of hospitalizations and ED visits, and probability of having a 30-day readmission. The finding of no discernable effects on costs and service utilization is consistent with the awardee's implementation experience and the study design challenges describe earlier. Nebraska Medicine reported difficulties with engaging patients, particularly the sickest and highest-risk patients. Thus, meeting some of the targeted outcomes might have been challenging, as these patients might have had more potential to improve their management of diabetes and avoid hospitalizations and ED visits. In addition, about one-quarter of enrollees dropped out of the program before completing the 90-day intervention period, which would likely dilute the program's impacts.

The program did, however, achieve its goal to increase primary care visits in ambulatory settings. According to the program's theory of action, by providing early and timely post-

discharge health coaching and RPM, the awardee sought to help patients improve diabetes-related self-care following hospital discharge, enabling them to stabilize their diabetes. As a result, patients would work more closely with their PCPs to continue their diabetes care. This finding is consistent with the opinions of RIISCC staff about the effects of the program on care delivery and patients' satisfaction. Health coaches observed that participating in the program encouraged patients to reengage with their PCPs. Coaches also reported seeing improved behavioral changes among their patients, such as increased exercise and improved diet.

Main findings from impact evaluation

- The RIISCC intervention was associated with an increase in primary care visits in ambulatory settings.
- The evidence suggests that the program had little or no effect on total Medicare expenditures, number of hospitalizations and ED visits, and the probability of having a 30-day readmission.

Two program components in particular might have contributed to the observed success in increasing primary care visits. First, the awardee provided consistent services to participants, achieved high levels of satisfaction among those who did not drop out from the program, and engaged local providers. Second, the awardee provided the staff and training needed to deliver the services. Although the awardee encountered struggles with staff retention, it improved its approach to staffing and operations during the course of the cooperative agreement. The health coaches and medical assistants received ongoing education from biweekly meetings with the program's medical director and the diabetes center's clinical director, working through examples of caring for program participants.

CONCLUSION

Findings from this impact analysis suggest that the RIISCC program of Nebraska Medicine achieved its goal to increase regular contact with primary care physicians, but likely did not reduce total Medicare expenditures, number of hospitalizations, or probability of having a 30-day readmission. The program also expected to reduce ED visits, but this study estimates that the program *increased* the proportion of patients having an ED visit by about 7 percentage points, and had no effect on the number of ED visits. These findings, together with results from earlier programs, suggest that interventions involving RPM and patient activation have the potential to help patients with diabetes better manage their own care by working closely with PCPs. However, for changes in diabetes management to affect the cost and use of care, it is likely to require a longer intervention and additional efforts to engage patients most likely to benefit from the intervention, such as those with higher risk and worse health status.

Limitations of evaluation

The key limitation of the analysis is its observational design, combined with the low participation rate among eligible beneficiaries. The study estimated program impacts by

comparing outcomes among participants to those of the matched comparison group. However, participants differed from eligible nonparticipants on one critical measure (mortality) not captured in pre-enrollment Medicare administrative files and claims, indicating clear selection bias. The comparison group matched the participant group well on observed measures but could not fully account for this unobservable selection bias. The mortality rate among participants was roughly half the mortality rate among eligible nonparticipants and much lower than that of comparison group members, suggesting the impact estimates would be biased. After controlling for mortality, the estimated (favorable) program impact on costs fell by about 40 percent, from a reduction of 16 percent to a reduction of 11 percent, and the favorable estimated effect on hospitalizations dropped to essentially zero. This raises the additional concern that surviving participants and comparison group members might have differed on other factors that were important determinants of costs but were not captured by claims data, leading to potential overestimates of the true program impacts.

PROGRAM SUSTAINABILITY

When its award ended in February 2018, Nebraska Medicine sustained its RIISCC program, although the program lacked sufficient funding to retain the entire scope of the program for new enrollees. New enrollees continue to receive the health coaching component and the RPM, but these and other aspects of the program are now more embedded with the PCP. To help sustain the program, the awardee obtained partial funding from shared-savings payments from Nebraska Medicine's accountable care organization. The awardee was also in the process of obtaining FFS payments from commercial payers for RPM and exploring options for billing Medicare and Medicaid for this service.

Nebraska Medicine's proposed payment model

The awardee proposed a bundled payment model, in which payers would reimburse the awardee a set amount per patient for one episode of services (defined as the 90-day RIISCC intervention). The payment would have covered the work of the entire team, rather than paying separately for the services each individual team member provided. The payer would then either pay the awardee more if it spent less than the target, or recover payment if it exceeded the target. The awardee drew from Medicare's Bundled Payments for Care Improvement initiative in designing its model.

By the end of the award period, Nebraska Medicine had not implemented its payment model, but continued to pursue it. Nebraska Medicine estimated a per patient per day cost of $17 to $22 from its own data, but had not yet estimated payer-specific costs from claims data. Lack of claims data from payers and a good comparison group hindered the awardee's ability to calculate program savings, which are key to generating payers' interest. Still, the awardee continued to assess the optimal time period for RPM services and how to provide them more efficiently, which could help sustain the health coaching component.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays rely on a difference-in-differences approach with beneficiary-level fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay, emergency department (ED) visit, or 30-day readmission is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay, any ED visit, and any 30-day readmission during the baseline period. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries).

A pseudo-enrollment date was assigned to each comparison beneficiary and eligible nonparticipant to define the end of the baseline period and start of the follow-up period, based on the date of the anchor encounter The anchor encounter was an inpatient admission or outpatient visit (ED or other outpatient visit) to one of the program hospitals (for eligible nonparticipants) or other Omaha hospitals (for comparison beneficiaries). The pseudo-enrollment date was equal to on the discharge date of the anchor encounter plus a random number of days selected to ensure that the distribution of days between the anchor encounter and pseudo-enrollment matched the distribution of days between anchor encounter and actual enrollment for participants.

Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of core outcomes used for this evaluation.

The impact analysis included only 23 percent of total participants, according to the awardee's data (Table A.1). The analysis identified participants based on the awardee's finder file. It dropped most of the excluded participants from the study because they were not Medicare beneficiaries or could not be found in the Medicare enrollment database through names-based matching (70 percent). It dropped another estimated 5 percent because they did not meet the study's standard claims-based inclusion criteria. The analysis did not include the remaining 2 percent of enrollees because they died within 30 days of enrollment, did not have a valid anchor encounter, or were considered an outlier based on their baseline characteristics.

Table A.1. Identification of the final sample for impact analysis for NMC

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through August 31, 2017		**1,903**
Not enrolled in Medicare or not identified in Medicare enrollment files[a]	1,346	557
Did not meet standard claims-based inclusion criteria:		
Not enrolled in both Medicare Parts A and B	66	491
Enrolled in Medicare Advantage	9	482
Medicare was not the primary payer	9	473
Lacked 90 days of FFS enrollment during baseline period	10	463
Died within 30 days of enrollment	4	459
No valid anchor encounter	27	432
Outlier during propensity score matching	2	430
Final analytic sample		**430**

Sources: Mathematica's analysis of information from the awardee's finder file from December 1, 2014, through August 31, 2017, and Medicare claims and enrollment data from September 2013 through February 2018, as of March 13, 2019.

[a] The awardee did not provide patient identifiers, so Medicare beneficiaries were identified through matching a participant's name, gender, and date of birth with information available from the Medicare enrollment database.

FFS = fee-for-service. NMC = University of Nebraska Medical Center.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. The table displays the weighted means of baseline characteristics for the 430 treatment beneficiaries and the 1,855 matched comparison beneficiaries used in the impact analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable, which was calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The analysis used four measures for exact matching: age category, sex, anchor encounter type, and dual Medicare–Medicaid eligibility. It then included a number of beneficiary characteristics as covariates in the propensity score model, including demographic characteristics (age, gender, and race); Medicare entitlement; health status (as measured by the hierarchical condition category [HCC] score and chronic condition indicators); quarter of enrollment; Medicare expenditures in total and by type of service; and service use. The analysis required an exact match on age category, sex, anchor encounter type, and dual Medicare–Medicaid eligibility. The analysis measured variables over various specified intervals within the 12 months before enrollment in the intervention. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

The table also shows the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence of the statistical significance of the difference in the means. The equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values equivalence test, which assesses whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the analysis also performed an omnibus test in which the null hypothesis is that the treatment and matched comparison groups are balanced across all linear combinations of the covariates. It used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

Table B.1. Baseline characteristics of treatment and matched comparison groups for NMC

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	68 (0.48)	68 (0.23)	-0.7 (0.69)	-1.1	< 0.01	0.29	< 0.01
White, %	78 (2.00)	77 (0.94)	0.6 (2.83)	0.8	0.01	0.83	< 0.01
Medicare entitlement and dual eligibility status, %							
Original reason for Medicare entitlement: disability	43 (2.39)	40 (1.12)	2.1 (3.41)	4.9	0.04	0.54	< 0.01
Health status and diagnosis							
HCC score[a]	2.0 (0.06)	2.0 (0.03)	0.0 (0.10)	-1.4	< 0.01	0.78	< 0.01
Acute renal failure, %	19 (1.88)	15 (0.75)	3.7 (2.52)	20	0.09	0.14	< 0.01
CHF, %	26 (2.11)	28 (1.00)	-2.4 (3.09)	-9.2	< 0.01	0.45	< 0.01
COPD, %	26 (2.11)	24 (0.97)	1.3 (2.92)	5.1	0.03	0.65	< 0.01
Diabetes with acute complications, %	2.8 (0.80)	2.5 (0.34)	0.3 (1.08)	11	0.02	0.77	< 0.01
Diabetes with chronic complications, %	63 (2.33)	63 (1.13)	-0.2 (3.13)	-0.3	< 0.01	0.95	< 0.01
Hyperlipidemia, %	88 (1.57)	88 (0.74)	-0.5 (2.27)	-0.6	< 0.01	0.81	< 0.01
Hypertension, %	93 (1.19)	95 (0.52)	-1.7 (1.58)	-1.8	< 0.01	0.29	< 0.01
Major depression, %	15 (1.71)	13 (0.79)	1.6 (2.43)	11	0.04	0.51	< 0.01
Morbid obesity, %	16 (1.76)	17 (0.81)	-1.2 (2.59)	-7.5	< 0.01	0.65	< 0.01
Specified heart arrhythmias, %	26 (2.12)	28 (1.03)	-2.1 (3.05)	-8.1	< 0.01	0.49	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Vascular disease, %	19 (1.91)	20 (0.91)	-0.3 (2.71)	-1.6	< 0.01	0.91	< 0.01
Enrollment quarter, %							
Quarter 3	0.2 (0.23)	0.2 (0.12)	0.0 (0.33)	0.0	< 0.01	1.00	< 0.01
Quarter 4	8.4 (1.34)	7.5 (0.59)	0.8 (1.82)	10	0.03	0.64	< 0.01
Quarter 5	6.0 (1.15)	7.5 (0.56)	-1.4 (1.74)	-24	< 0.01	0.41	< 0.01
Quarter 6	6.0 (1.15)	5.8 (0.49)	0.2 (1.61)	3.5	< 0.01	0.89	< 0.01
Quarter 7	10 (1.48)	10 (0.70)	0.1 (2.14)	0.9	< 0.01	0.97	< 0.01
Quarter 8	15 (1.72)	15 (0.84)	0.0 (2.48)	0.2	< 0.01	0.99	< 0.01
Quarter 9	16 (1.75)	16 (0.87)	-0.1 (2.49)	-0.9	< 0.01	0.96	< 0.01
Quarter 10	12 (1.55)	11 (0.73)	0.5 (2.22)	4.5	0.02	0.81	< 0.01
Quarter 11	6.3 (1.17)	7.2 (0.56)	-0.9 (1.67)	-15	< 0.01	0.58	< 0.01
Quarter 12	11 (1.51)	11 (0.76)	-0.2 (2.10)	-1.5	< 0.01	0.94	< 0.01
Quarter 13	7.4 (1.27)	5.8 (0.57)	1.6 (1.65)	22	0.06	0.33	< 0.01
Quarter 14	2.1 (0.69)	2.8 (0.40)	-0.7 (1.05)	-31	< 0.01	0.53	< 0.01
Medicare expenditures							
Total expenditures, 12 months before enrollment	19,741 (1262.38)	19,564 (455.09)	177 (1851.45)	0.9	< 0.01	0.92	< 0.01
Total expenditures, 3 months before enrollment	9,476 (596.69)	8,559 (201.48)	917 (813.90)	9.7	0.07	0.26	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total expenditures, 3 months before enrollment[b]	9,199 (535.43)	8,442 (195.79)	757 (754.68)	8.2	0.07	0.32	< 0.01
Total payment amount, day of enrollment ($ PBPM)	3,243 (891.21)	2,777 (183.30)	465 (1191.81)	14	0.03	0.70	< 0.01
Total payment amount, day of enrollment ($ PBPM)[b]	2,296 (554.59)	2,148 (127.72)	148 (792.80)	6.5	0.01	0.85	< 0.01
Service use							
Total hospitalizations, 12 months before enrollment, per beneficiary	0.9 (0.07)	0.9 (0.02)	0.0 (0.10)	-1.7	< 0.01	0.88	< 0.01
Total ED or observation visits, 12 months before enrollment, per beneficiary	1.2 (0.12)	1.1 (0.04)	0.1 (0.15)	11	0.06	0.37	< 0.01
Any ED or observation visit, 30 days before enrollment, %	34 (2.28)	31 (1.08)	2.7 (3.22)	7.9	0.06	0.41	< 0.01
Any IP admission, 30 days before enrollment, %	35 (2.30)	34 (1.03)	0.7 (3.12)	1.9	0.01	0.83	< 0.01
Propensity Score	0.0 (< 0.01)	0.0 (< 0.01)	0.0 (< 0.01)	11	0.11	0.06	< 0.01
Number of beneficiaries	430	1,855					
Omnibus test				Chi-squared statistic 172.997	Degrees of freedom 51.000	P-value 0.000	

Source: Mathematica's analysis of Medicare claims data.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The analysis calculated the comparison group means in the table by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid. Exact matching variables include age category, sex, anchor encounter type, and dual Medicare–Medicaid eligibility.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; HCC = hierarchical condition category; IP = inpatient; NMC = University of Nebraska Medical Center; PBPM = per beneficiary per month; SE = standard error.

Appendix C

Detailed results from impact estimates

This page has been left blank for double-sided copying.

Table C.1. Estimated impact of the NMC intervention on select Medicare FFS expenditures ($ PBPM) and use measures during a 12-month follow-up period, top coded, without controlling for mortality

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)[b]					
Baseline year	1,539	1,529			
Months 1–6	1,659	1,912	-263* (140)	-14%	0.06
Months 7–12	1,403	1,601	-208 (131)	-13%	0.11
Months 1–12	1,558	1,840	-291** (123)	-16%	0.02
Acute inpatient expenditures ($ PBPM)					
Baseline	689	614			
Months 1–6	677	717	-115 (89)	-15%	0.19
Months 7–12	539	580	-117 (80)	-19%	0.14
Months 1–12	614	688	-150* (79)	-20%	0.06
Outpatient expenditures ($ PBPM)					
Baseline	329	309			
Months 1-6	315	347	-52 (32)	-15%	0.10
Months 7-12	301	300	-20 (31)	-6.7%	0.52
Months 1-12	304	330	-47 (32)	-14%	0.14
Professional Part B expenditures ($ PBPM)					
Baseline	327	350			
Months 1–6	335	380	-21 (18)	-6.1%	0.24
Months 7–12	298	340	-19 (18)	-6.2%	0.30
Months 1–12	316	369	-30* (16)	-8.7%	0.07
Home health expenditures ($ PBPM)					
Baseline	62	64			
Months 1–6	100	95	7.3 (14)	8.1%	0.59
Months 7–12	65	77	-9.5 (11)	-13%	0.41
Months 1–12	88	88	2.5 (11)	2.9%	0.83
SNF expenditures ($ PBPM)					
Baseline	56	53			
Months 1–6	95	112	-19 (23)	-18%	0.41
Months 7–12	103	108	-6.7 (23)	-6.7%	0.77
Months 1–12	108	127	-21 (23)	-17%	0.36
Durable medical equipment expenditures ($ PBPM)					
Baseline	36	33			
Months 1–6	45	37	4.9 (4.0)	12%	0.23
Months 7–12	46	33	9.9** (4.4)	29%	0.02
Months 1–12	47	37	6.5* (3.9)	17%	0.10

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Acute hospital admissions, per 1,000 beneficiaries					
Baseline	825	875			
Months 1–6	718	829	-61 (83)	-8.0%	0.47
Months 7–12	642	756	-64 (80)	-9.4%	0.42
Months 1–12	675	795	-70 (71)	-9.3%	0.33
Outpatient ED or observation visits, per 1,000 beneficiaries					
Baseline	1,370	1,095			
Months 1–6	1,234	986	-28 (113)	-2.2%	0.80
Months 7–12	1,171	935	-39 (119)	-3.3%	0.74
Months 1–12	1,211	964	-28 (96)	-2.2%	0.77
Primary care visits in any settings, per 1,000 beneficiaries					
Baseline	9,081	8,813			
Months 1–6	9,789	9,198	322 (454)	3.4%	0.48
Months 7–12	8,973	8,678	26 (439)	< 1%	0.95
Months 1–12	9,505	9,057	180 (405)	1.9%	0.66
Primary care visits in ambulatory settings, per 1,000 beneficiaries					
Baseline	6,803	6,525			
Months 1–6	7,747	6,848	621** (311)	8.8%	0.05
Months 7–12	7,170	6,530	361 (316)	5.4%	0.25
Months 1–12	7,534	6,721	535* (277)	7.7%	0.05
Specialist visits in any setting, per 1,000 beneficiaries					
Baseline	13,437	14,212			
Months 1–6	14,701	15,250	227 (679)	1.6%	0.74
Months 7–12	13,141	13,812	104 (690)	< 1%	0.88
Months 1–12	14,058	14,906	-74 (627)	< 1%	0.91
Specialist visits in ambulatory settings, per 1,000 beneficiaries					
Baseline	9,510	9,720			
Months 1–6	10,517	10,311	417 (445)	4.2%	0.35
Months 7–12	9,636	9,073	774* (455)	8.8%	0.09
Months 1–12	10,067	9,734	544 (411)	5.8%	0.19
Percentage of participants with a hospital admission					
Baseline	50	50			
Months 1–6	26	26	-0.13 (2.3)	< 1%	0.96
Months 7–12	24	24	0.40 (2.4)	1.7%	0.87
Months 1–12	40	38	1.9 (2.6)	5.0%	0.46
Percentage of participants with an outpatient ED visit					
Baseline	62	52			
Months 1–6	33	29	3.7 (2.5)	13%	0.14

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 7–12	32	28	4.4* (2.6)	16%	0.10
Months 1–12	50	43	6.9** (2.7)	16%	0.01
Percentage of participants with a 30-day readmission					
Baseline	4.8	6.1			
Months 1–6	8.0	10	-2.1 (1.5)	-21%	0.17
Months 7–12	3.3	6.3	-3.0** (1.2)	-47%	0.02
Months 1–12	10	13	-2.9* (1.7)	-22%	0.09

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicare claims and enrollment data as of August 10, 2018.

Note: Impact estimates for number of visits or stays rely on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; NMC = University of Nebraska Medical Center; PBPM = per beneficiary per month; SE = standard error; SNF = skilled nursing facility.

Table C.2. Estimated impact of the NMC intervention on select Medicare FFS expenditures ($ PBPM) and use measures during a 12-month follow-up period, top coded, controlling for mortality

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)[b]					
Baseline year	1,539	1,529			
Months 1-6	1,685	1,840	-164 (138)	-9.2%	0.23
Months 7-12	1,436	1,548	-122 (129)	-8.3%	0.34
Months 1-12	1,620	1,796	-186 (120)	-11%	0.12
Acute inpatient expenditures ($ PBPM)					
Baseline	689	614			
Months 1-6	695	667	-48 (86)	-7.0%	0.58
Months 7-12	555	546	-67 (79)	-12%	0.40
Months 1-12	656	660	-80 (76)	-12%	0.30
Outpatient expenditures ($ PBPM)					
Baseline	329	309			
Months 1-6	316	343	-47 (32)	-14%	0.14
Months 7-12	301	297	-17 (31)	-5.9%	0.58
Months 1-12	307	328	-42 (32)	-12%	0.19
Professional Part B expenditures ($ PBPM)					
Baseline	327	350			
Months 1-6	337	373	-12 (18)	-3.5%	0.51
Months 7-12	302	335	-10 (18)	-3.4%	0.58
Months 1-12	322	365	-20 (16)	-5.9%	0.23
Home health expenditures ($ PBPM)					
Baseline	62	64			
Months 1-6	101	92	10 (14)	12%	0.44
Months 7-12	65	76	-9.0 (11)	-12%	0.43
Months 1-12	90	87	4.7 (11)	5.6%	0.68
SNF expenditures ($ PBPM)					
Baseline	56	53			
Months 1–6	96	109	-16 (23)	-16%	0.50
Months 7–12	103	105	-4.7 (23)	-4.8%	0.84
Months 1–12	109	125	-18 (23)	-15%	0.43
Durable medical equipment expenditures ($ PBPM)					
Baseline	36	33			
Months 1–6	45	37	4.7 (4.3)	12%	0.27
Months 7–12	47	33	9.9** (4.4)	29%	0.03
Months 1–12	47	37	6.4 (4.1)	16%	0.12

Table C.2 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Acute hospital admissions, per 1,000 beneficiaries					
Baseline	825	875			
Months 1–6	733	783	-0.47 (82)	< 1%	1.00
Months 7–12	661	725	-13 (79)	-2.1%	0.86
Months 1–12	713	773	-9.9 (69)	-1.4%	0.89
Outpatient ED or observation visits, per 1,000 beneficiaries					
Baseline	1,370	1,095			
Months 1–6	1,241	965	0.22 (115)	< 1%	1.00
Months 7–12	1,170	922	-28 (120)	-2.3%	0.82
Months 1–12	1,225	954	-5.0 (97)	< 1%	0.96
Primary care visits in any settings, per 1,000 beneficiaries					
Baseline	9,081	8,813			
Months 1–6	9,831	9,082	481 (460)	5.2%	0.30
Months 7–12	9,045	8,596	181 (440)	2.1%	0.68
Months 1–12	9,608	8,990	350 (408)	3.9%	0.39
Primary care visits in ambulatory settings, per 1,000 beneficiaries					
Baseline	6,803	6,525			
Months 1–6	7,752	6,841	632** (311)	8.9%	0.04
Months 7–12	7,170	6,525	367 (317)	5.4%	0.25
Months 1–12	7,537	6,714	545** (277)	7.8%	0.05
Specialist visits in any setting, per 1,000 beneficiaries					
Baseline	13,437	14,212			
Months 1–6	14,812	14,903	683 (668)	5.0%	0.31
Months 7–12	13,279	13,583	471 (686)	3.8%	0.49
Months 1–12	14,350	14,732	392 (617)	2.9%	0.53
Specialist visits in ambulatory settings, per 1,000 beneficiaries					
Baseline	9,510	9,720			
Months 1–6	10,514	10,294	431 (447)	4.3%	0.33
Months 7–12	9,637	9,080	767* (455)	8.7%	0.09
Months 1–12	10,085	9,740	556 (411)	5.9%	0.18
Percentage of participants with a hospital admission					
Baseline	50	50			
Months 1–6	26	24	1.3 (2.3)	5.5%	0.55
Months 7–12	24	23	1.1 (2.4)	4.7%	0.65
Months 1–12	40	36	3.3 (2.5)	8.9%	0.19
Percentage of participants with an outpatient ED visit					
Baseline	62	52			
Months 1–6	33	28	4.2* (2.5)	15%	0.10

Table C.2 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 7–12	32	28	4.6* (2.6)	17%	0.08
Months 1–12	50	43	7.4*** (2.7)	17%	< 0.01
Percentage of participants with a 30-day readmission					
Baseline	4.8	6.1			
Months 1–6	8.0	9.5	-1.5 (1.5)	-16%	0.32
Months 7–12	3.3	6.0	-2.7** (1.2)	-45%	0.03
Months 1–12	10	13	-2.2 (1.7)	-17%	0.20

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicare claims and enrollment data as of August 10, 2018.

Note: Impact estimates for number of visits or stays rely on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; NMC = University of Nebraska Medical Center; PBPM = per beneficiary per month; SE = standard error; SNF = skilled nursing facility.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, analysis also estimated the program impacts for the University of Nebraska Medicine Center (NMC) using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in Round 1 of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to NMC. The analysis calculated probabilities using the results of a Bayesian regression that jointly models impacts on CMS's four core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for CMS's four core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. All NMC estimates included in the Bayesian analysis control for mortality. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for NMC led to a Bayesian estimate of the program's impact on total Medicare expenditures of -5 percent (an estimated reduction of $87 per beneficiary per month) in the first year.

Table D.1. Comparison of frequentist and Bayesian impact estimates for NMC in the first year after enrollment

	Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	-186 (-421, 49)	-87 (-237, 55)	-2%	-11%	-5%
Hospital admissions	-9.9 (-146, 126)	-33 (-93, 27)	-2%	-1%	-5%
ED visits	-5.0 (-196, 186)	-62 (-165, 34)	-3%	> -1%	-5%
Readmissions	-2.2 (-5.5, 1.1)	-0.6 (-1.6, 0.4)	-2%	-17%	-5%

Source: Mathematica's analysis of information from Medicare claims and enrollment data from September 2014 through August 2017, as of March 13, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: Total expenditures include both Medicare Parts A and B spending. Readmissions are the percentage of beneficiaries with a readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions rely on data from the HCIA R1 evaluation.

Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

HCIA R1= Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month.

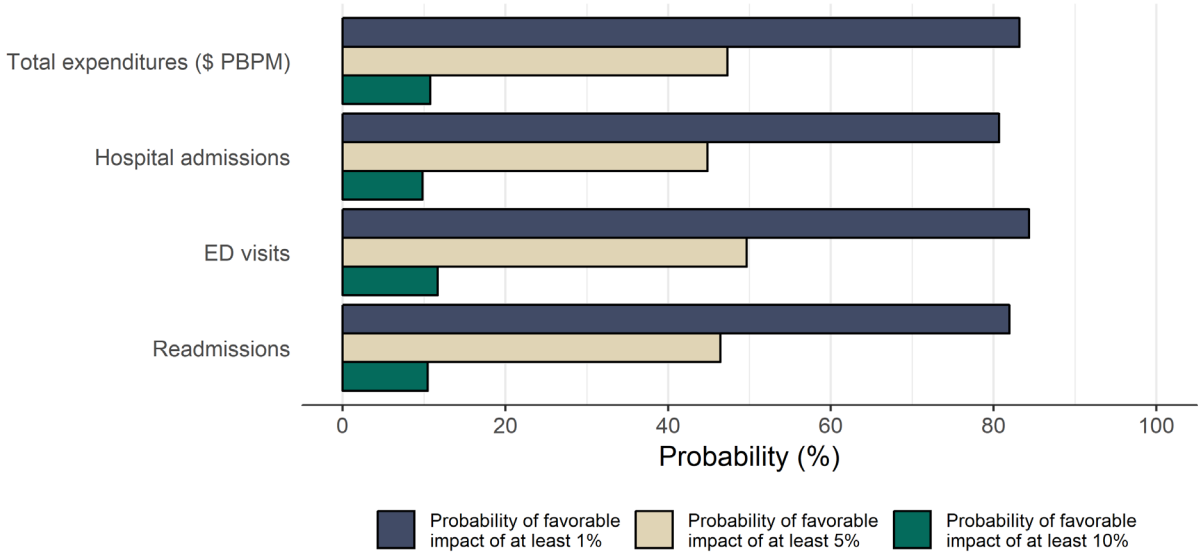
Because the frequentist results relied on a small sample and are therefore imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates that are more consistent across outcomes. The Bayesian impact estimates are so consistent across outcomes because the Bayesian model borrows strength across outcomes, drawing all four estimates closer to the average percentage impact across outcomes for this awardee. The highly favorable

(although statistically insignificant) percentage impacts for total Medicare expenditures and readmissions drive this average, leading to unexpectedly favorable Bayesian estimates for hospitalizations and ED visits.

Although the magnitudes of the Bayesian and frequentist impact estimates differ substantially – for example, the Bayesian analysis increases the percentage impact on hospital admissions from -1 percent to -5 percent – the two sets of results are not incompatible. Despite more favorable point estimates and narrowed uncertainty intervals, none of the Bayesian impact estimates is statistically distinguishable from zero, corroborating the results of the frequentist analysis. At the same time, the uncertainty intervals for the frequentist and Bayesian impact estimates overlap substantially, indicating that the two sets of estimates are not statistically distinguishable from each other.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that NMC achieved favorable impacts in the first year on CMS's four core outcomes at three different thresholds: a favorable impact of (1) 1 percent or more, (2) 5 percent or more, and (3) 10 percent or more.

Figure D.1. Probability that the NMC program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from Medicare claims and enrollment data from September 2014 through August 2017, as of March 13, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: Total expenditures include both Medicare Parts A and B spending. Readmissions are the percentage of beneficiaries with a readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions rely on data from the HCIA R1 evaluation.

HCIA R1= Round 1 of the Health Care Innovation Awards; NMC = University of Nebraska Medical Center.

There is a high probability—more than 80 percent—that NMC had a favorable impact of 1 percent or more on total Medicare expenditures, hospital admissions, ED visits, and readmissions. In fact, there is more than a 40 percent probability that NMC had a favorable impact of 5 percent or more on all three outcomes. The likelihood of a 10 percent or larger impact is small (about 1 in 10). These probabilities indicate that it is possible that the program had moderately favorable impacts. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the NMC program may have slightly favorable effects, but the evidence does not indicate meaningful reduction in total expenditures or service use.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: New York City Health + Hospitals, New York

September 2020

Anna Hill, Rivka Weiser, Stefanie Pietras, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

NEW YORK CITY HEALTH + HOSPITALS, NEW YORK

New York City Health + Hospitals (NYC H+H), a public benefit corporation that serves as the public safety net in the city's health care system, used funding it received through Round 2 of the Health Care Innovation Awards (HCIA R2) to develop and support the Emergency Department (ED) Care Management Initiative at 6 of its 11 hospitals. The goals of the initiative were to improve follow-up care among patients with ambulatory care-sensitive conditions (ACSCs) who visited the ED and reduce their use of acute care. To be eligible for the program, a patient who visited the ED at a participating site must have been able to be discharged from the ED safely and have (1) visited the ED for an ACSC; (2) met other service use criteria (for example, had another recent ED visit or hospitalization); or (3) been deemed likely to benefit from the program based on the clinical judgment of a nurse care manager (NCM) or referring ED clinician. The program provided care management in the ED, along with 90 days of follow-up care coordination. The program began in September 2014 and ended in August 2017. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program provided care management and 90-day care coordination to eligible patients who visited the ED in six NYC H+H hospitals. It aimed to help patients better manage their health and avoid unnecessary hospitalizations and repeated visits to the ED, thereby lowering the cost of their care.
- The analysis compared 45,277 Medicaid beneficiaries and 9,134 Medicare FFS beneficiaries who had an ED visit for an ACSC at a participating NYC H+H hospital with Medicaid and Medicare beneficiaries with similar demographic and health characteristics who had an ED visit for an ACSC at NYC hospitals that did not participate in the program.
- Because it was not possible to replicate the program's selection criteria using claims, the evaluation used an intent-to-treat (ITT) design. The evaluation selected beneficiaries for the analytic sample who met the sample selection criteria regardless of program participation status. As a result, the analytic sample included 24 percent of Medicaid beneficiaries and 30 percent of Medicare beneficiaries identified as participants by the awardee, but only 18 percent of the Medicare and 16 percent of the Medicaid claims-based ITT analytic samples.

The awardee hypothesized that providing interdisciplinary care management and extended care coordination would help ED patients with ACSCs manage their health and avoid unnecessary hospitalizations and repeated visits to the ED, thereby lowering the cost of their care. An ED-based care team at each site (including NCMs, community liaison workers [CLWs], pharmacists, and home care intake nurses) helped patients manage their health by providing education, support, and links to ambulatory care and home health care as needed. The goals of the program were to reduce 30-day hospitalizations by 35 percent, reduce 7- and 30-day repeat ED visits by 25 percent, and produce a net savings among program participants of $75 million.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The NYC H+H ED Care Management Initiative provided care management in the ED and 90-day care coordination (which covered the transition to comprehensive ambulatory care) to eligible patients who visited the ED.
Major innovation	An ED-based interdisciplinary team of NCMs, CLWs, pharmacists, and home care intake nurses helped patients better manage their health by providing education, support, and links to ambulatory care and home health care as needed.
Program components	• Care management during an ED visit, including performing a risk assessment, creating an ambulatory care plan, providing health education and counseling as needed, referring participants to primary care providers and/or specialists as appropriate, and referring appropriate participants to the home care intake nurse and/or pharmacist (who were also part of the program staff). • Transitional care coordination for up to 90 days after ED discharge included contact with patients to encourage and remind them to keep appointments for follow-up care.
Target population	Adult patients were eligible for the program if they met any of the following requirements and could be safely discharged from the ED: • Visited the ED for ACSCs • Visited the ED and met other service use criteria • Were deemed likely to benefit from the program based on the clinical judgment of the NCM or referring ED clinician
Participating providers	Six EDs at NYC H+H Hospitals in New York City
Total enrollment	The awardee reported enrolling 83,946 patients, representing about 84 percent of its original target.
Level of engagement	The awardee reported that, although it was in general easy to engage participants in the ED, it was challenging to reach and engage them by telephone after their discharge. In addition, the extent to which program staff provided the intended 90-day transitional care coordination varied, based on staff capacity.
Theory of change or theory of action	NYC H+H hypothesized that providing care management and extended care coordination would help ED patients with ACSCs and other patients whose conditions were high-risk or high-cost to better manage their health and avoid unnecessary hospitalizations and repeated visits to the ED, thereby lowering the cost of their care.
Award amount	$17,916,663
Effective launch date	September 2014, immediately after award
Program settings	Six NYC H+H EDs
Market area	Urban, New York City
Target outcomes	• Reduce 30-day hospitalizations by 35 percent • Reduce 7- and 30-day repeat ED visits by 25 percent • Produce net savings of $75 million among program participants

Table 1 (*continued*)

Program characteristics	Description
Payment model	NYC H+H proposed three models to pay for the ED Care Management Initiative: (1) a model that incorporated the program in existing global risk-capitated contracts with two Medicaid and Medicare managed care plans, (2) a value-based payment model that would have adjusted Medicaid and commercial FFS payments for quality performance, and (3) a shared savings model with an all-payer ACO. As of August 2018, NYC H+H had not pursued any of these models.
Sustainability plans	NYC H+H had originally planned to sustain and scale the ED Care Management Initiative program to all of its 11 hospitals by implementing ED Care Triage, a similar program under a new DSRIP program. As of August 2018, however, the awardee had not done this but was using DSRIP funds to maintain and augment aspects of the program at some of its sites.

ACO = accountable care organization; ACSC = ambulatory care-sensitive condition; CLW = community liaison worker; DSRIP = Delivery System Reform Incentive Payment; ED = emergency department; FFS = fee-for-service; NCM = nurse care manager; NYC H+H = New York City Health and Hospitals.

The impact analysis presented in this report includes 45,277 Medicaid beneficiaries and 9,134 Medicare fee-for-service (FFS) beneficiaries who had an ED visit with a primary or secondary diagnosis of an ACSC at a participating NYC H+H hospital and met other evaluation selection criteria. The comparison group included 47,602 Medicaid and 9,901 Medicare FFS beneficiaries with similar demographic and health characteristics who had an ED visit for an ACSC at hospitals in New York City that did not participate in the intervention and who also met the same evaluation selection criteria. Because it was not possible to replicate the program's selection criteria using claims, the evaluation selected beneficiaries for the analytic sample who met the claims-based sample inclusion criteria regardless of program participation status. Table 2 summarizes the key features of the impact evaluation and the identification of the Medicare and Medicaid analytic samples. Appendix A, Tables A.1 and A.2 provide detailed information on the identification of the Medicare and Medicaid analytic samples.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group. Because it was not possible to replicate the program's selection criteria using claims, the evaluation selected beneficiaries for the analytic sample who met the sample inclusion criteria regardless of program participation status.
Intervention group for evaluation	The awardee reported total enrollment of nearly 84,000 participants, of whom 35,493 could be linked to Medicare or Medicaid enrollment data. The eligible treatment group for this impact analysis included 9,134 Medicare FFS beneficiaries and 45,277 Medicaid beneficiaries who had an ED visit with a primary or secondary diagnosis of an ACSC at a participating NYC H+H hospital from September 1, 2014, to February 28, 2017, and met the other study inclusion criteria. The group included only 24 percent of 30,105 Medicaid beneficiaries and 30 percent of 5,388 Medicare beneficiaries identified as participants by the awardee who could be linked to Medicare or Medicaid enrollment data. After applying the claims- and enrollment-based study inclusion criteria, the analytic sample included 1,627 Medicare and 7,140 Medicaid beneficiaries identified by the awardee as participants.

Table 2 (*continued*)

Features	Description
Intervention group for evaluation (*continued*)	The study dropped most of the excluded participants because they did not have a qualifying ED visit within two days of the reported enrollment date (3,861 Medicaid participants and 2,229 Medicare participants) or because there was no primary or secondary diagnosis of an ACSC associated with the ED visit (8,623 Medicaid participants and 1,202 Medicare participants). It dropped nearly one-third of the Medicaid participant sample because of lack of enrollment in Medicaid at the time of the ED visit or because of dual eligibility status (9,000 Medicaid participants). The Medicare sample included the dually eligible beneficiaries. Another large group of Medicaid participants were not enrolled in Medicaid for at least 90 days during the 12-month baseline period (1,385). The study excluded 9 Medicaid participants because they died within 30 days of the ED visit and 87 because they did not live in New York. It dropped 330 Medicare participants because of lack of Medicare FFS enrollment in the 90 days before or at the time of the ED visit, enrollment in Medicare Advantage, death within 30 days of the ED visit, or residency outside of New York.
Comparison group	The comparison group included 47,602 Medicaid and 9,901 Medicare beneficiaries with similar demographic and health characteristics who had an ED visit for an ACSC at NYC hospitals that did not participate in the program.
Limitations	Because it was not possible to replicate the program's selection criteria using claims, the study selected beneficiaries for the analytic sample who met the sample selection criteria regardless of program participation status. The low participation rate in the analytic sample (about 16 to 18 percent) would have made it difficult to identify impacts if measured over all eligible beneficiaries. If participants differed from eligible nonparticipants in ways not captured in Medicaid or Medicare administrative files and claims, the impact estimates might be biased.

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; NYC H+H = New York City Health and Hospitals.

PROGRAM DESIGN AND ADAPTATION

The ED Care Management Initiative service delivery model had two components: care management and transitional care coordination.[1] Because the awardee tailored the program to local needs and capacity, the roles and responsibilities of the NCM and CLW staff between the two components varied from one hospital to another.

Care management

In general, NCMs used data in electronic medical records or spoke with on-duty ED clinicians to identify potential participants when the patient was still in the ED. After recruiting patients and enrolling them in the program, the NCMs provided initial care management services before discharge from the ED. Their services included performing a risk assessment, creating an ambulatory care plan, providing health education and counseling, referring participants to primary care providers and/or specialists, and referring appropriate participants to the home care intake nurse and/or pharmacist (who were also part of the program staff). Upon referral, the home care intake nurse assessed the participant's need for home care and linked him or her to appropriate services. A staff pharmacist also provided services to patients enrolled in the care management component. Pharmacy-related services included educating participants about

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the ED Care Management Initiative. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

medications that might help them manage their conditions and helping them access necessary medications. Finally, the NCMs checked in with the participant by telephone within 24 to 48 hours after discharge from the ED to assess any short-term needs.

Transitional care coordination after ED discharge

CLWs provided longer-term transitional care coordination services to discharged patients. They checked in with participants by telephone at 30, 60, and 90 days after enrollment (the intervention ended at 90 days). Depending on patients' needs, the CLWs were available to remind participants of their upcoming ambulatory care visits, follow up with patients after these visits, and link patients to other providers and resources, including other care team staff.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

By the end of the program, the awardee had enrolled more than 83,000 participants (about 84 percent of its enrollment target). Initially, the awardee had a difficult time meeting its program enrollment target. Therefore, it focused on increasing its enrollment beginning in the first year, which led to enrollment of patients unlikely to benefit from the program and increased workload challenges for program staff. In January 2017, program leadership also revisited the eligibility criteria, agreeing that criteria should be more flexible and allow for greater clinical judgment to identify those expected to benefit most from the program (such as high-acuity patients, frequent ED users, and those with social and behavioral health issues).

Implications of program implementation for detecting impacts

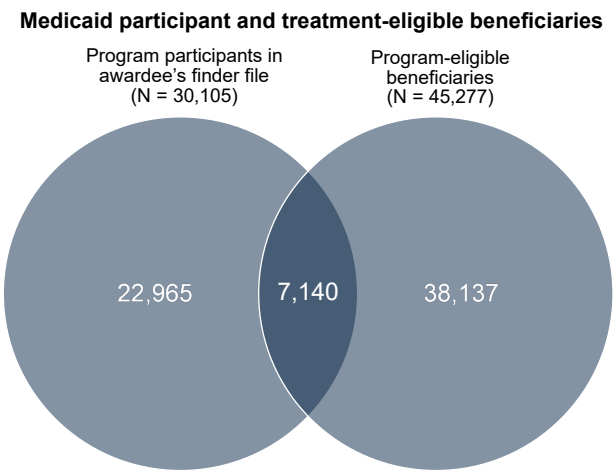
- Many factors (such as pressure to increase enrollment, demanding workloads, primary care capacity issues, limited resources to address social determinants of health, and difficulty with ongoing enrollee engagement) challenged the awardee's ability to provide comprehensive care management and transitional care coordination effectively.
- To be eligible for the program, a prospective enrollee had to have visited the ED for an ACSC and/or met other use thresholds (for example, a recent hospitalization or ED visit). The program considered additional factors, such as history of substance abuse or not having a designated primary care physician (PCP), in determining a person's appropriateness for enrollment.
- During most of the program, program staffs' broad targeting of ED patients for enrollment (not all of whom would necessarily benefit from the program) might have diluted the impact of the program on patients' outcomes.

Pressure to increase enrollment, demanding workloads, and staffing vacancies challenged the awardee's ability to provide comprehensive care management and transitional care coordination. In interviews, program staff noted that heavy emphasis on enrollment meant the NCMs had less time to manage comprehensive care and coordinate transitional care. Staff thought the quality of their services suffered as a result. For example, instead of the follow-up calls at 30, 60,

and 90 days in the original intervention design, some sites completed only some calls or prioritized calls for participants with certain conditions because of staffing and time constraints. The program also faced challenges in standardizing processes across sites, although it worked to address these issues (for example, by establishing cross-site work groups to share best practices). In addition, according to program leaders and staff, lack of staff to address enrollees' social and behavioral determinants of health (including homelessness and substance abuse) led to problems engaging patients and hindered the program's effectiveness. Therefore, although it was relatively easy to engage patients in the ED, it often was difficult to reach and engage them by telephone after their discharge. Although some sites had social workers in the ED (not funded by the cooperative agreement), the awardee acknowledged that its intervention did not have resources available to identify and meet these needs. Finally, although the program tried to connect enrollees to appropriate ambulatory care (such as PCPs), limited PCP capacity in the NYC H+H system challenged staff's ability to secure prompt ambulatory care appointments. Despite staffing challenges, participating clinicians and other implementation staff perceived the program as having a positive effect on the delivery of care.

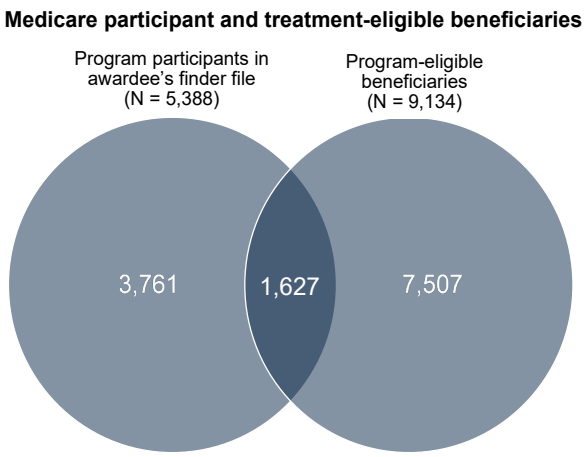
ESTIMATING PROGRAM IMPACTS

Selecting treatment and comparison samples



Because it was not possible to replicate all the program's enrollment criteria using claims (for example, clinical judgment), the study used an ITT design, which measures the impact of the ED Care Management Initiative on program-eligible Medicare and Medicaid beneficiaries. That is, the study drew the sample from a group of beneficiaries who visited participating EDs during the implementation period (September 2014 through February 2017), regardless of program participation status.[2] The group consisted of Medicare and Medicaid beneficiaries who met evaluation eligibility criteria and visited a participating ED for an ACSC (the program selection criterion observable in claims). The study drew the matched comparison sample from beneficiaries who went to EDs in New York City that did not participate in the program and met the same eligibility criteria for the evaluation.

[2] The analytic sample is restricted to those with a qualifying ED visit from September 1, 2014, to February 28, 2017, even though the program ended in August 2017. This allows for the 90-day period of the care management program plus an additional three-month follow-up period during which the participant might continue to receive program services if there were delays in care during the initial 90-day period.



A total of 45,277 Medicaid and 9,134 Medicare beneficiaries met the analytic sample selection criteria. Of the 83,946 participants the awardee reported, the finder file it provided included 30,105 Medicaid beneficiaries and 5,388 Medicare beneficiaries. Of the 45,277 eligible Medicaid beneficiaries identified using claims, 7,140 (16 percent) participated in the ED Care Management Initiative. Of the 9,134 eligible Medicare beneficiaries identified using Medicare claims, 1,627 (18 percent) participated in the ED Care Management Initiative.

The study dropped most of the excluded participants because they did not have a qualifying ED visit within two days of the reported enrollment date or because there was no primary or secondary diagnosis of an ACSC associated with the ED visit (Appendix A, Tables A.1 and A.2). Moreover, the awardee reported that it was not unusual for no claim to be submitted for ED visits at their facilities, so it is not surprising that many beneficiaries were found to have no ED visits. The study dropped nearly one-third of the Medicaid participant sample because of lack of enrollment in Medicaid at the time of the index ED visit or because of dual eligibility status (the Medicare sample included the dually eligible beneficiaries). Another large group of Medicaid participants were not enrolled in Medicaid for at least 90 days during the 12-month baseline period.

Characteristics of treatment and comparison group beneficiaries

Comparing treatment and comparison group characteristics at baseline confirmed that the two groups were well balanced (Table 3). Appendix B provides the full balance results, including all characteristics measured during the 12 months before enrollment. Treatment-eligible beneficiaries appeared to be slightly sicker and more costly than comparison beneficiaries, although all beneficiaries in the analysis were at high risk for using services and high expenditures. Among Medicaid beneficiaries, the average Chronic Illness and Disability Payment System (CDPS) score was 2.3 and 2.4 for the treatment and comparison groups, respectively. Among Medicare beneficiaries, the average hierarchical condition category (HCC) score was about 1.6 for both the treatment and comparison groups, indicating that beneficiaries in the analytic sample were likely to have 60 percent higher health expenditures in the subsequent year than the typical Medicare beneficiary (average HCC score of 1).

All beneficiaries selected for the analytic sample had an ED visit with a primary or secondary diagnosis of an ACSC in the baseline period per the sample inclusion criteria, and the follow-up period began on the day following the ED visit triggering eligibility. Medicaid beneficiaries had higher use of acute care in the baseline period than did Medicare beneficiaries. For example, 37

percent of treatment-eligible Medicare beneficiaries and 33 percent of comparison beneficiaries had at least three ED visits in the baseline period, compared with 46 percent of the Medicaid treatment-eligible sample and 44 percent of the Medicaid comparison sample.

Table 3. Selected baseline characteristics of treatment and comparison group beneficiaries

Measure	Medicaid beneficiaries		Medicare beneficiaries	
	Treatment eligible (N = 45,277)	Comparison (N = 47,602)	Treatment eligible (N = 9,134)	Comparison (N = 9,901)
Demographics				
Age at enrollment, years	44	42	68	68
Female, %	64	58	57	60
Dually eligible for Medicare and Medicaid, %	NA	NA	67	67
Risk score				
HCC[a]	n.a.	n.a.	1.57	1.63
CDPS score[b]	2.28	2.37	n.a.	n.a.
Service use and expenditures during the year before enrollment				
Number of hospitalizations (per 1,000 beneficiaries)	576	481	788	693
Number of ED or observation visits (per 1,000 beneficiaries)[e]	2,859	2,605	2,479	2,263
Number of ED or observation visits for an ACSC (per 1,000 beneficiaries)[c]	1,098	1,028	666	643
Total Medicare expenditures ($ PBPM)	NA	NA	2,217	2,064

Source: Mathematica's analysis of awardee-provided enrollment data and Medicare and Medicaid FFS claims, encounter, and enrollment data from September 2012 to February 2018.

Notes: The baseline period covers the 12-month period before the ED visit that initiated selection in the analytic sample.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare or Medicaid. In addition to the number of months enrolled in FFS Medicare or Medicaid, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

Appendix B presents full balance results, including all beneficiary characteristics observed at baseline. All standardized differences between mean treatment and comparison characteristics in this table were below 0.1, except for age among Medicaid beneficiaries. The standardized difference in average age between treatment and comparison Medicaid beneficiaries was 0.13.

The Medicare analysis required an exact match for the quarter of enrollment and ACSC diagnosis. The Medicaid analysis required an exact match for the quarter of enrollment, ACSC diagnosis, and county (New York City borough) of residency.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of 1 represents average expected expenditures. The evaluation calculated HCC scores by using the most recently available HCC algorithms.

[b] The CDPS score relied on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

[c] The number of outpatient ED visits is top-coded at the 98th percentile of the treatment group distribution for both groups because of the presence of more extreme outliers in the treatment group than in the comparison group.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; HCC = hierarchical condition category; n.a. = not applicable; NA = not available; PBPM = per beneficiary per month.

Analytic approach

The difference-in-differences study design compares the change in outcomes between the baseline and follow-up periods among treatment-eligible beneficiaries to the change during the same period among the matched comparison beneficiaries. Difference-in-differences models produce unbiased impact estimates when the treatment and comparison groups have parallel trends in outcomes before the implementation of the intervention. Selecting a matched comparison group ensures balance across key patient and facility characteristics likely to affect expenditure and use outcomes. The study defined the baseline period as the year before (and including the day of) each beneficiary's ED visit for an ACSC, and it defined the follow-up period as the year after the visit. The primary outcomes were total Medicare spending, number of hospital admissions, and number of ED visits. Because 90 percent of the Medicaid sample was enrolled in the Comprehensive Managed Care Program with no observed expenditures in Medicaid claims, this analysis did not include any Medicaid expenditure outcomes. The study estimated impacts separately for Medicare and Medicaid beneficiaries. Appendix A describes the statistical models used to estimate the effects of the program and the sample selection criteria in detail.

IMPACT RESULTS

Overall, the impact estimates of the ED Care Management Initiative did not provide evidence that the program affected any of the main outcomes among all Medicaid beneficiaries. However, there is evidence of a favorable impact on ED use among Medicaid beneficiaries who became eligible for the program during the first nine months of implementation. The rate of ED visits declined by 8.2 percent among Medicaid beneficiaries eligible in the first nine months of the program (Table 4) and the proportion of beneficiaries who had an all-cause ED visit also declined by an estimated 6.5 percent. There is also evidence that all-cause hospitalizations declined by 6.3 percent among treatment beneficiaries who became eligible during the first nine months of the program.

There are marked differences in impact estimates using the group of Medicaid beneficiaries who became eligible for the program after the first nine months relative to beneficiaries who became eligible within the first nine months. In particular, the rate of all-cause hospitalizations increased by 7.3 percent among treatment beneficiaries, which is opposite and nearly equal in magnitude to the favorable estimate observed among Medicaid beneficiaries who became eligible within the first nine months.

Table 4. Estimated impact of the ED Care Management Initiative on selected outcomes during a 12-month follow-up period among Medicaid beneficiaries

	All Medicaid beneficiaries	Medicaid beneficiaries meeting eligibility criteria in first nine months of program	Medicaid beneficiaries meeting eligibility criteria after the first nine months of program
Number of ED or observation visits (per 1,000 beneficiaries, per year)[a]			
Impact (count)	-27	-213***	22
Percentage impact	-1.2%	-8.20%	1.0%
p-value	0.31	< 0.01	0.44
Number of ED or observation visits for an ACSC (per 1,000 beneficiaries, per year)[a]			
Impact (count)	-15	-74***	0.05
Percentage impact	-2.7%	-10%	< 1%
p-value	0.31	< 0.01	1.00
Percentage of beneficiaries with any ED or observation visits			
Impact (pp)	-1.5	-4.3***	-1.0
Percentage impact	-2.4%	-6.50%	-1.6%
p-value	0.15	< 0.01	0.35
Number of hospitalizations (per 1,000 beneficiaries, per year)			
Impact (count)	22	-57*	43***
Percentage impact	3.3%	-6.30%	7.3%
p-value	0.11	0.09	< 0.01
Sample sizes			
Treatment	45,277	9,747	35,530
Comparison	47,602	9,995	37,607

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicaid FFS claims, encounter, and enrollment data from September 2012 to February 2018.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between baseline and intervention periods. The impact estimates for the binary outcomes of any ED visit are regression-adjusted treatment–comparison differences based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having more than two visits at baseline. The intervention months or years are beneficiary specific and defined relative to each beneficiary's date of enrollment or index ED visit. Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate. Appendix C shows full impact estimates. Appendix D shows the results of the Bayesian analysis.

[a] Because of more extreme outliers in the treatment group than in the comparison group for the number of ED visits, these estimates of impacts on the number of all-cause ED visits and the number of ED visits for ACSCs were obtained after trimming these outcomes for both groups at the 98th percentile of the treatment group distribution.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; pp = percentage point.

Among Medicare beneficiaries, there were few statistically significant impact estimates. However, the proportion of Medicare beneficiaries who had an all-cause ED visit declined by an estimated 8.8 percent, and ED use was one of the initiative's target outcomes. The rate of ED visits also declined, but this result was not statistically significant (Table 5).

The results of this analysis provide little support for concluding that the ED Care Management Initiative reduced total Medicare expenditures. The reduction in total Medicare expenditures after one year of follow-up was less than 1 percent. Consistent with a lack of impact on expenditures, the impact estimate for the number of annualized, all-cause hospitalizations was small and not distinguishable from zero.

Finally, the ED Care Management Initiative reduced the number of specialist visits in the 12-month follow-up period among both Medicare and Medicaid beneficiaries (see Appendix C). Because the program had goals of both connecting enrollees to appropriate services (in some cases, specialists) and reducing unnecessary use of services, it is difficult to know whether this reduction is a favorable impact of the program. Moreover, the awardee's theory of change provided no rationale for reduced specialist visits due to the program. (Appendix D presents the results from the Bayesian analysis.)

Table 5. Estimated impact of the ED Care Management Initiative on selected outcomes during a 12-month follow-up period among Medicare beneficiaries

	All Medicare beneficiaries
Number of ED or observation visits (per 1,000 beneficiaries, per year)[a]	
Impact (count)	-55
Percentage impact	-3.10%
p-value	0.18
Number of ED or observation visits for an ACSC (per 1,000 beneficiaries, per year)[a]	
Impact (count)	-14
Percentage impact	-3.80%
p-value	0.43
Percentage of beneficiaries with any ED or observation visits	
Impact (pp)	-5.3**
Percentage impact	-8.80%
p-value	0.03
Total expenditures ($ PBPM)	
Impact ($)	-13
Percentage impact	< 1%
p-value	0.88
Number of hospitalizations (per 1,000 beneficiaries, per year)	
Impact (count)	-14
Percentage impact	-1.50%
p-value	0.65

Table 5 (*continued*)

	All Medicare beneficiaries
Sample sizes	
Treatment	9,134
Comparison	9,901

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS claims and enrollment data from September 2012 to February 2018.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between baseline and intervention periods. The impact estimates for the binary outcomes of any ED visit are regression-adjusted treatment–comparison differences based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having more than two visits at baseline. The intervention months or years are beneficiary specific and defined relative to each beneficiary's date of enrollment or index ED visit. Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate. Appendix C shows full impact estimates. Appendix D shows the results of the Bayesian analysis.

[a] Because of more extreme outliers in the treatment group than in the comparison group for the number of ED visits, these estimates of impacts on the number of all-cause ED visits and the number of ED visits for ACSCs were obtained after trimming these outcomes for both groups at the 98th percentile of the treatment group distribution.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; pp = percentage point.

The modest impact estimates found in this analysis are consistent with both the awardee's implementation experience and the study design challenges described earlier. First, operational challenges were likely to diminish the impact of the program. Staff reported difficulty providing enrolled beneficiaries with consistent and successful care coordination and care management services because of internal NCM and CLW staffing and workload issues, as well as systemic problems with ambulatory care provider capacity. They also reported difficulty engaging patients after discharge from the ED, even when programs applied all aspects of the ED Care Management Initiative as planned. This was, in part, because of the high prevalence of social issues (such as homelessness), which made it difficult to coordinate care management and transitional care. Second, broad outreach to ED patients, not all of whom might benefit from care management and 90 days of care coordination, to improve enrollment in the program likely diluted the estimated impact of the program on patients' outcomes. Program leaders and staff noted that some patients who met the official inclusion criteria, such as those with urinary tract infections (UTIs; an ACSC), did not need the

Main findings from impact evaluation

- The ED Care Management Initiative led to meaningful reductions in ED visits, mainly among high-risk Medicaid beneficiaries.
- There was little estimated effect of the program on total Medicare expenditures and use of inpatient services.
- Modest impacts are consistent with operational challenges that program staff faced, including staffing shortages, difficulty engaging patients after discharge from the ED, and limited PCP capacity in the NYC H+H system.

services the program offered. Finally, the analytic sample contained many beneficiaries who did not participate in the intervention (more than 80 percent of the treatment-eligible group did not participate), which would further dilute estimates of impacts.

Favorable impact findings on the use of ED services among Medicaid beneficiaries eligible within the first nine months of the intervention suggests that the intervention might be most effective when it is provided to patients who are in poorer health status and are very frequent users of the ED. This is consistent with the awardee's theory of change, which suggests that extended care management could help reduce repeated ED visits and costs for patients at high risk. Beneficiaries who became eligible for the program during the first nine months of implementation appeared to be at higher risk of frequent use of acute care: they had significantly higher average CDPS scores and higher baseline use of acute care than the group who became eligible later in the implementation period. Differences in acuity between the two groups could be due to changes in eligibility criteria made at the end of the first year, which sought to increase enrollment, and could have resulted in including more lower-risk participants during the latter portion of the implementation period. However, the study did not observe a similar pattern among Medicare beneficiaries. Thus, it is plausible that the observed Medicaid differences were due to unobserved factors that differed over time for the hospitals from which the study drew eligible patients, and which have nothing to do with the program.

CONCLUSION

The ED Care Management Initiative led to meaningful reductions in all-cause ED visits among high-risk Medicaid beneficiaries who became eligible for the program within its first 9 months, and a reduction in the likelihood of frequent ED use among both Medicare and high-risk Medicaid beneficiaries. The program had little estimated effect on total Medicare expenditures and use of inpatient services. Although there was evidence that the program reduced specialist visits in the 12 months following program enrollment, it is unclear that these reductions represented favorable outcomes of the program.

Program staff and leaders suggested that more careful outreach to enroll patients in the intervention could improve program impacts in two ways. First, patients most likely to benefit would receive the intervention, and those unlikely to benefit (for example, those with diagnoses of UTIs) would no longer be enrolled. Second, more restrictive outreach would result in fewer enrolled beneficiaries, which could help to alleviate the staffing and workload issues that NCMs and CLWs reported. It might also be easier to find prompt primary care for a smaller group of patients. Care management of ED patients is unlikely to be effective without accessible follow-up ambulatory care following discharge. Finally, the awardee acknowledged the difficulty of providing ongoing care to patients affected by social and behavioral health issues, which are prevalent in the high-risk group seeking care at NYC H+H EDs.

Limitations of evaluation

This analysis had several limitations. Because it was not possible to replicate the selection process in claims, the analytic sample included patients with diagnoses that would qualify them for inclusion in the intervention, regardless of participation status, which might reduce the observed impact of the program. The analysis also did not include beneficiaries who went to the ED without a primary or secondary diagnosis of an ACSC but were enrolled based on a clinician's judgment and who might have differed from beneficiaries included in the analytic sample. As a result, of the 45,277 treatment-eligible Medicaid beneficiaries using Medicaid enrollment and claims data, 7,140 (16 percent) participated in the ED Care Management Initiative. Similarly, of the 9,134 eligible Medicare beneficiaries identified using Medicare claims, only 1,627 (18 percent) participated in the Initiative. Finally, it is not possible to rule out differences between treatment and comparison groups in the extent of regression toward the mean. Because the matching process relied on a single year of baseline data, it is possible that the treatment-eligible and comparison samples were well matched at only a single point in time, but had different underlying time paths of the key outcomes that could affect the impact estimates found in a difference-in-differences analysis. Diagnostic checks for regression toward the mean using two-year baseline data suggest that this potential problem is of limited concern.

PROGRAM SUSTAINABILITY

After its award ended in August 2017, the awardee continued work similar to that of the ED Care Management Initiative using funding under New York's Delivery System Reform Incentive Payment (DSRIP) program. It continued to fund the positions of some program staff, specifically NCMs and CLWs, at some participating sites, although these staff no longer received oversight from program leaders or had to provide the standard set of program services. The Initiative also piloted a new but similar program (called ED Care Triage) at one of the implementing sites, which used the same staff from the HCIA R2 program but featured more formal tracking and standardized services to meet the milestones and metrics required under the DSRIP.

Leaders at NYC H+H reported implementing two other initiatives after the award ended, which had goals similar to those of the ED Care Management Initiative. First, in early 2018, the awardee opened an urgent care clinic (ExpressCare) managed by the ED and that redirects certain patients, particularly those who do not have urgent or resource-intensive needs, to more appropriate care settings. Second, in May 2018, the awardee began to

NYC H+H's proposed payment model

NYC H+H proposed three models to pay for the ED Care Management Initiative. The first model would have incorporated the program in existing global risk-capitated contracts with two Medicaid and Medicare managed care plans. The second option involved a value-based payment model that would have adjusted Medicaid and commercial FFS payments for quality performance. The awardee also considered a shared savings model with an all-payer accountable care organization. However, as of August 2018, the awardee had not pursued any of these models.

roll out a systemwide care management program, which features NCM and social worker dyads who provide care management services in the ED, inpatient, and ambulatory settings.

NYC H+H initially proposed three payment models: (1) a global risk-capitated model, (2) a value-based payment model, and (3) a shared savings model. As of August 2018, however, the awardee had not pursued any of these models. Instead, it used DSRIP funding, as noted earlier, to continue the work of some former HCIA R2 care management staff and to pilot the ED Care Triage program. DSRIP goals and broader institutional goals also had motivated NYC H+H to focus on managing care systemwide and diverting some people, as appropriate, from the ED to urgent care (via ExpressCare). Program leaders noted that the systemwide care management program and ExpressCare were for qualifying patients, regardless of payer.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach with beneficiary-level fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay, any emergency department (ED) visit, and more than two ED visits is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or any ED visit during the baseline period. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date, for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard core outcomes used for this evaluation.

This impact analysis was an intent-to-treat (ITT) analysis, which means it estimated the impact of the ED Care Management Initiative on a group of treatment-eligible beneficiaries. Because it was not possible to replicate selection into the program using Medicare claims, an ITT analysis reduces the risk of selection bias in the impact estimates. The study drew the analytic sample from a group of Medicare fee-for-service (FFS) and Medicaid beneficiaries who had an ED visit for an ambulatory care-sensitive condition (ACSC) at a participating site during the implementation period (Tables A.1 and A.2). It was further limited to beneficiaries who were enrolled in Medicare FFS (both Parts A and B, with Medicare as the primary payer) or Medicaid when their eligibility for awardee-provided services began (that is, their enrollment date). Beneficiaries also had to meet all of these program criteria for at least 90 days during the baseline year (the 365 days immediately before their enrollment plus the date of their enrollment ED visit). The study excluded beneficiaries who died within 30 days of enrollment.

Table A.1. Identifying the final sample for impact analysis for NYC H+H, Medicare

	Number of participants removed from analytic sample	Numer of participants remaining in analytic sample
Total number of program participants in awardee's finder file		**5,388**
No qualifying ED visit	2,229	3,159
No diagnosis of ACSC	1,202	1,957
Participants who did not meet the standard claims-based inclusion criteria		
Not enrolled in Medicare FFS Parts A and B at time of ED visit	149	1,808
Enrolled in Medicare Advantage	15	1,793
Medicare was not the primary payer	18	1,775
Lacked 90 days of enrollment during baseline period	90	1,685
Died within 30 days of enrollment	4	1,681
State of residency is not New York	54	1,627
Number of participants in analytic sample		**1,627**
Number of nonparticipant treatment-eligible beneficiaries in sample		7,507
Final analytic sample		**9,134**

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS claims and enrollment data from September 2012 to February 2018.

Note: 605 beneficiaries from the finder file could not be linked to either Medicare or Medicaid enrollment data, and they are excluded from the count in this row.

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; NYC H+H = New York City Health + Hospitals.

Table A.2. Identifying the final sample for impact analysis for NYC H+H, Medicaid

	Number of participants removed from analytic sample	Numer of participants remaining in analytic sample
Total number of program participants in awardee's finder file		**30,105**
No qualifying ED visit	3,861	26,244
No diagnosis of ACSC	8,623	17,621
Participants who did not meet the standard claims-based inclusion criteria		
Not enrolled in Medicaid at time of ED visit[a]	9,000	8,621
Lacked 90 days of enrollment during baseline period	1,385	7,236
Died within 30 days of enrollment	9	7,227
State of residency is not New York	87	7,140
Number of participants in analytic sample		**7,140**
Number of nonparticipant treatment-eligible beneficiaries in sample		38,137
Final analytic sample		**45,277**

Table A.2 (*continued*)

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicaid FFS claims, encounter, and enrollment data from September 2012 to February 2018.

Note: 605 beneficiaries from the finder file could not be linked to either Medicare or Medicaid enrollment data, and they are excluded from the count in this row.

[a] The Medicare sample includes beneficiaries who were dually eligible for Medicare and Medicaid and, therefore, excludes them from the Medicaid sample. The count in this row also includes Medicaid beneficiaries with restricted benefits in their enrollment month.

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; NYC H+H = New York City Health + Hospitals.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Tables B.1 and B.2 show the results from matching for the Medicare and Medicaid analyses, respectively. The tables display the weighted means of baseline characteristics for the treatment and matched comparison beneficiaries used in the impact analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable. The study calculated the standardized difference as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit.

The table also shows the results of the equivalency-of-means tests. The *p*-values come from a weighted two-sample t-test, which provides evidence of a statistically significant difference in the means. The equivalence test *p*-values are the greater of the two one-sided weighted t-test *p*-values equivalence test, which assesses whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the study conducted an omnibus test in which the null hypothesis was that the treatment and matched comparison groups balanced across all linear combinations of the covariates. The study used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

For the Medicare analysis, the matching variables included demographic characteristics (age, gender, and race); county (New York City borough) of residency; local area demographics (race, ethnicity, poverty, education level, and number of primary care providers per 100,000 residents); hospital characteristics (number of beds, disproportionate share hospital share payment, and occupancy rate); Medicare entitlement and dual eligibility status; primary or secondary ambulatory care-sensitive condition (ACSC) diagnosis; health status (as measured by the hierarchical condition category score and selected chronic condition indicators); Medicare expenditures in total and by type of service; and service use. The study required an exact match for the quarter of enrollment and ACSC diagnosis. It measured the variables over specified intervals within the 12 months before enrollment in the intervention.

For the Medicaid analysis, the matching variables included demographic characteristics (age, gender, and race); hospital characteristics (number of beds, disproportionate share hospital share payment, occupancy rate); Medicare entitlement and dual eligibility status; primary or secondary ACSC diagnosis; health status (as measured by the Chronic Illness and Disability Payment System score and selected chronic condition indicators); and service use. The study required an exact match for the quarter of enrollment, ACSC diagnosis, and county (New York City borough) of residency. It measured the variables over specified intervals within the 12 months before enrollment in the intervention.

For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

Table B.1. Baseline characteristics of treatment and matched comparison groups for NYC H+H, Medicare sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	68 (0.16)	68 (0.15)	-0.71 (0.23)	-1.1	-0.05	< 0.01	< 0.01
Female, %	57 (0.52)	60 (0.49)	-2.4 (0.73)	-4.2	-0.05	< 0.01	< 0.01
White, %	27 (0.47)	28 (0.49)	-0.42 (0.64)	-1.5	-0.01	0.52	< 0.01
Black, %	48 (0.52)	48 (0.48)	0.55 (0.74)	1.1	0.01	0.46	< 0.01
Other, %	10.0 (0.31)	9.9 (0.33)	0.11 (0.43)	1.1	0.00	0.80	< 0.01
Hispanic, %	12 (0.34)	13 (0.33)	-0.60 (0.49)	-4.9	-0.02	0.22	< 0.01
Unknown, %	2.3 (0.16)	2.0 (0.15)	0.36 (0.21)	16	0.02	0.08	< 0.01
Medicare entitlement and dual eligibility status, %							
Dually eligible for Medicare and Medicaid	67 (0.49)	67 (0.47)	-0.30 (0.69)	< +/-1	-0.01	0.67	< 0.01
Original reason for Medicare entitlement: disability	42 (0.52)	41 (0.49)	1.6 (0.74)	3.8	0.03	0.03	< 0.01
Health status and diagnoses							
HCC score[a]	1.57 (0.01)	1.63 (0.01)	-0.06 (0.02)	-3.7	-0.04	< 0.01	< 0.01
Vascular disease, %	23 (0.44)	24 (0.45)	-1.2 (0.61)	-5.3	-0.03	0.05	< 0.01
Morbid obesity, %	6.1 (0.25)	7.7 (0.25)	-1.6 (0.36)	-27	-0.07	< 0.01	< 0.01
Major depression, %	11 (0.33)	11 (0.33)	-0.04 (0.47)	< +/-1	0.00	0.93	< 0.01
COPD, %	14 (0.36)	15 (0.37)	-1.7 (0.52)	-12	-0.05	< 0.01	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Diabetes with acute complications, %	1.5 (0.13)	1.6 (0.11)	-0.05 (0.18)	-3.6	0.00	0.76	< 0.01
Congestive heart failure, %	17 (0.39)	21 (0.41)	-4.1 (0.59)	-25	-0.11	< 0.01	< 0.01
Medicare expenditures							
Total expenditures	2,217 (45)	2,064 (32)	152 (58)	6.9	0.04	< 0.01	< 0.01
Total expenditures[b]	2,088 (36)	1,968 (28)	121 (49)	5.8	0.04	0.01	< 0.01
Total expenditures, 3 months before enrollment	2,648 (65)	2,513 (48)	135 (86)	5.1	0.02	0.11	< 0.01
Total expenditures, 3 months before enrollment[b]	2,426 (49)	2,325 (39)	102 (66)	4.2	0.02	0.12	< 0.01
Total expenditures, 7 days before enrollment[b]	4,830 (137)	5,142 (130)	-312 (204)	-6.5	-0.02	0.13	< 0.01
Physician services expenditures	335 (5.7)	377 (5.3)	-41 (8.4)	-12	-0.08	< 0.01	< 0.01
Physician services expenditures[b]	313 (4.2)	346 (4.0)	-33 (6.2)	-11	-0.08	< 0.01	< 0.01
Outpatient expenditures	357 (16)	361 (7.4)	-3.9 (18)	-1.1	0.00	0.83	< 0.01
Outpatient expenditures[b]	324 (6.6)	332 (5.9)	-8.9 (9.4)	-2.7	-0.01	0.34	< 0.01
Home health expenditures	60 (2.2)	70 (2.4)	-9.9 (3.3)	-16	-0.05	< 0.01	< 0.01
Home health expenditures, 3 months before enrollment	71 (3.4)	76 (3.4)	-4.9 (4.9)	-6.9	-0.02	0.31	< 0.01
Service use							
Total hospitalizations	788 (22)	693 (13)	95 (25)	12	0.05	< 0.01	< 0.01
Total hospitalizations (top-coded)[b]	685 (14)	659 (11)	26 (19)	3.8	0.02	0.17	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total hospitalizations, 3 months before enrollment	953 (31)	894 (21)	59 (39)	6.2	0.02	0.13	< 0.01
Total hospitalizations, 3 months before enrollment[b]	799 (21)	822 (18)	-23 (29)	-2.9	-0.01	0.42	< 0.01
Total ED or observation visits	2,849 (62)	2,348 (28)	500 (72)	18	0.08	< 0.01	< 0.01
Total ED or observation visits (top-coded)[b]	2,479 (28)	2,263 (19)	216 (37)	8.7	0.08	< 0.01	< 0.01
Total ED or observation visits, 3 months before enrollment	6,215 (92)	5,418 (41)	797 (107)	13	0.09	< 0.01	< 0.01
Total ED or observation visits, 3 months before enrollment (top-coded)[b]	5,707 (42)	5,246 (27)	460 (55)	8.1	0.11	< 0.01	< 0.01
Primary care visits, any setting	6,780 (105)	7,469 (89)	-688 (146)	-10	-0.07	< 0.01	< 0.01
Primary care visits, any setting (top-coded)[b]	6,393 (81)	7,107 (73)	-714 (115)	-11	-0.09	< 0.01	< 0.01
Primary care visits, any setting, 3 months before enrollment	7,523 (134)	8,607 (111)	-1,084 (188)	-14	-0.08	< 0.01	< 0.01
Primary care visits, any setting, 3 months before enrollment (top-coded)[b]	6,978 (100)	8,121 (89)	-1,142 (143)	-16	-0.12	< 0.01	< 0.01
Primary care visits, ambulatory setting	5,255 (80)	5,425 (69)	-170 (109)	-3.2	-0.02	0.12	< 0.01
Primary care visits, ambulatory setting (top-coded)[b]	4,966 (62)	5,181 (59)	-215 (87)	-4.3	-0.04	0.01	< 0.01
Specialist visits, ambulatory setting	10,099 (128)	10,682 (129)	-583 (182)	-5.8	-0.05	< 0.01	< 0.01
Specialist visits, ambulatory setting (top-coded)[b]	9,764 (111)	10,353 (112)	-589 (159)	-6.0	-0.06	< 0.01	< 0.01
Specialist visits, all settings	15,031 (200)	14,766 (165)	265 (263)	1.8	0.01	0.31	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Primary ACSC diagnosis, %							
None	60 (0.51)	60 (0.49)	0.00 (0.71)	< +/-1	0.00	1.00	< 0.01
Angina	0.11 (0.03)	0.22 (0.04)	-0.11 (0.06)	-105	-0.03	0.06	< 0.01
Asthma	3.8 (0.20)	5.0 (0.19)	-1.2 (0.30)	-33	-0.06	< 0.01	< 0.01
Chest pain	9.2 (0.30)	11 (0.29)	-1.4 (0.43)	-15	-0.05	< 0.01	< 0.01
Cellulitis	3.1 (0.18)	1.9 (0.16)	1.1 (0.23)	37	0.07	< 0.01	< 0.01
COPD	1.1 (0.11)	1.1 (0.12)	0.04 (0.16)	3.3	0.00	0.81	< 0.01
Dehydration	1.1 (0.11)	1.1 (0.13)	0.00 (0.16)	< +/-1	0.00	1.00	< 0.01
Diabetes	3.9 (0.20)	3.9 (0.18)	-0.04 (0.28)	-1.1	0.00	0.88	< 0.01
DVT	0.10 (0.03)	0.22 (0.03)	-0.13 (0.06)	-128	-0.04	0.03	< 0.01
Heart failure	0.19 (0.05)	0.34 (0.07)	-0.15 (0.08)	-82	-0.04	0.04	< 0.01
Hypertension	4.8 (0.22)	4.4 (0.22)	0.48 (0.31)	9.8	0.02	0.13	< 0.01
Intractable pain	1.7 (0.13)	0.72 (0.08)	0.94 (0.16)	57	0.07	< 0.01	< 0.01
Pneumonia	1.1 (0.11)	1.2 (0.11)	-0.10 (0.16)	-8.9	-0.01	0.54	< 0.01
Seizure	2.1 (0.15)	1.6 (0.13)	0.49 (0.19)	23	0.03	< 0.01	< 0.01
Sickle cell	0.37 (0.06)	0.41 (0.04)	-0.03 (0.10)	-8.8	-0.01	0.73	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Syncope	3.1 (0.18)	2.4 (0.18)	0.67 (0.24)	22	0.04	< 0.01	< 0.01
UTI	4.7 (0.22)	5.3 (0.23)	-0.57 (0.32)	-12	-0.03	0.07	< 0.01
Secondary ACSC diagnosis, %							
None	40 (0.51)	40 (0.49)	0.00 (0.71)	< +/-1	0.00	1.00	< 0.01
Asthma	2.1 (0.15)	3.0 (0.17)	-0.91 (0.23)	-44	-0.06	< 0.01	< 0.01
Chest pain	2.9 (0.18)	2.6 (0.14)	0.33 (0.24)	11	0.02	0.17	< 0.01
Cellulitis	0.45 (0.07)	0.75 (0.08)	-0.30 (0.12)	-67	-0.05	< 0.01	< 0.01
COPD	0.97 (0.10)	1.1 (0.10)	-0.08 (0.14)	-8.4	-0.01	0.57	< 0.01
Dehydration	0.73 (0.09)	0.76 (0.09)	-0.02 (0.13)	-3.0	0.00	0.86	< 0.01
Diabetes	15 (0.37)	17 (0.36)	-2.1 (0.53)	-14	-0.06	< 0.01	< 0.01
DVT	0.20 (0.05)	0.07 (0.03)	0.13 (0.05)	64	0.03	0.02	< 0.01
Heart failure	0.48 (0.07)	0.37 (0.06)	0.11 (0.09)	24	0.02	0.22	< 0.01
Hypertension	30 (0.48)	28 (0.47)	1.5 (0.67)	5.1	0.03	0.02	< 0.01
Intractable pain	2.5 (0.16)	1.1 (0.09)	1.4 (0.19)	56	0.09	< 0.01	< 0.01
Pneumonia	0.27 (0.05)	0.17 (0.05)	0.10 (0.07)	38	0.02	0.13	< 0.01
Seizure	1.3 (0.12)	1.1 (0.10)	0.21 (0.16)	16	0.02	0.21	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Sickle cell	0.11 (0.03)	0.10 (0.03)	0.01 (0.05)	5.0	0.00	0.91	< 0.01
Syncope	0.96 (0.10)	0.90 (0.09)	0.07 (0.14)	6.8	0.01	0.64	< 0.01
UTI	2.1 (0.15)	2.6 (0.15)	-0.49 (0.22)	-23	-0.03	0.03	< 0.01
Non-ACSC primary diagnosis, %							
Abdominal pain	2.7 (0.17)	3.5 (0.16)	-0.77 (0.26)	-28	-0.05	< 0.01	< 0.01
Dizziness or vertigo	2.9 (0.18)	3.4 (0.18)	-0.50 (0.27)	-17	-0.03	0.06	< 0.01
Genitourinary symptoms	1.2 (0.11)	1.0 (0.10)	0.14 (0.16)	12	0.01	0.36	< 0.01
Headache; including migraine	1.3 (0.12)	1.9 (0.13)	-0.56 (0.19)	-41	-0.05	< 0.01	< 0.01
Injuries from external causes	2.7 (0.17)	1.8 (0.15)	0.95 (0.22)	35	0.06	< 0.01	< 0.01
Other connective tissue disease	2.9 (0.18)	3.7 (0.17)	-0.74 (0.26)	-25	-0.04	< 0.01	< 0.01
Other lower respiratory disease	3.3 (0.19)	4.0 (0.17)	-0.72 (0.27)	-22	-0.04	< 0.01	< 0.01
Other nontraumatic joint disorders	3.3 (0.19)	3.6 (0.18)	-0.33 (0.26)	-10	-0.02	0.20	< 0.01
Spondylosis, other back problems	3.2 (0.18)	2.3 (0.16)	0.88 (0.25)	28	0.05	< 0.01	< 0.01
Superficial injury, contusion	2.5 (0.16)	2.5 (0.18)	0.03 (0.24)	1.3	0.00	0.89	< 0.01
Adverse outcomes within 90 days of enrollment, %							
Died within 90 days of enrollment	1.1 (0.11)	1.6 (0.12)	-0.45 (0.17)	-39	-0.04	< 0.01	< 0.01
Hospice use in first quarter after enrollment	0.57 (0.08)	0.70 (0.08)	-0.13 (0.12)	-23	-0.02	0.25	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Hospitalization on day of enrollment	1.1 (0.11)	0.48 (0.08)	0.62 (0.13)	56	0.06	< 0.01	< 0.01
Propensity score	0.09 (0.00)	0.08 (0.00)	0.01 (0.00)	12	0.11	< 0.01	< 0.01
Number of beneficiaries	9,134	9,901					
Omnibus test				Chi-squared statistic 7892.17	Degrees of freedom 114.00	*p*-value 0.00	

Sources: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS claims and enrollment data from September 2012 to February 2018. Area-level factors from the Area Health Resource File 2016–-2017 release.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. p-values come from a weighted two-sample t-test; equivalence test p-values are the greater of the p-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table were calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include primary and secondary ACSC diagnoses and quarter of enrollment.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of 1 represents average expected expenditures. The study used the most recently available HCC algorithms to calculate HCC scores.

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

ACSC = ambulatory care-sensitive condition; COPD = chronic obstructive pulmonary disease; DVT = deep vein thrombosis; ED = emergency department, FFS = fee-for-service; HCC = hierarchical condition category; NYC H+H = New York City Health + Hospitals; SE = standard error; UTI = urinary tract infection.

Table B.2. Baseline characteristics of treatment and matched comparison groups for NYC H+H, Medicaid sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	42 (0.07)	44 (0.07)	-2.0 (0.10)	-4.9	-0.14	< 0.01	< 0.01
Male, %	42 (0.23)	36 (0.22)	5.9 (0.33)	14	0.12	< 0.01	< 0.01
White, %	1.1 (0.05)	1.2 (0.05)	-0.09 (0.07)	-8.5	-0.01	0.18	< 0.01
Black, %	4.5 (0.10)	4.1 (0.09)	0.40 (0.13)	8.8	0.02	< 0.01	< 0.01
Hispanic or Latino, %	23 (0.20)	24 (0.19)	-1.7 (0.28)	-7.6	-0.04	< 0.01	< 0.01
Asian, %	1.1 (0.05)	1.1 (0.05)	0.02 (0.07)	1.4	0.00	0.84	< 0.01
Other, %	0.13 (0.02)	0.19 (0.02)	-0.05 (0.03)	-38	-0.01	0.05	< 0.01
Unknown, %	71 (0.21)	69 (0.21)	1.4 (0.30)	2.1	0.03	< 0.01	< 0.01
Medicaid enrollment, %							
Enrolled in comprehensive managed care	90 (0.14)	89 (0.15)	1.5 (0.21)	1.7	0.05	< 0.01	< 0.01
HCBS waiver, %	0.64 (0.04)	0.61 (0.04)	0.03 (0.05)	5.2	0.00	0.53	< 0.01
Health status and diagnoses							
CDPS score[a]	2.2 (0.01)	2.3 (0.01)	-0.08 (0.01)	-3.7	-0.04	< 0.01	< 0.01
AIDS or other infectious disease, %	16 (0.17)	17 (0.16)	-1.4 (0.25)	-8.7	-0.04	< 0.01	< 0.01
Cancer, %	2.6 (0.07)	2.6 (0.07)	-0.02 (0.10)	< +/-1	0.00	0.83	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Cardiovascular disease, %	48 (0.23)	52 (0.23)	-4.0 (0.33)	-8.3	-0.08	< 0.01	< 0.01
Central nervous system condition, %	16 (0.17)	15 (0.16)	0.94 (0.24)	6.0	0.03	< 0.01	< 0.01
Cerebrovascular condition, %	1.4 (0.06)	1.6 (0.05)	-0.19 (0.08)	-13	-0.02	0.02	< 0.01
Developmental disability, %	1.0 (0.05)	0.79 (0.05)	0.23 (0.06)	23	0.02	< 0.01	< 0.01
Disabled, %	17 (0.18)	20 (0.17)	-2.7 (0.25)	-16	-0.07	< 0.01	< 0.01
Eye condition, %	5.9 (0.11)	7.1 (0.11)	-1.3 (0.16)	-22	-0.05	< 0.01	< 0.01
Foster child, %	0.12 (0.02)	0.14 (0.02)	-0.02 (0.02)	-16	-0.01	0.42	< 0.01
Gastrointestinal condition, %	17 (0.18)	19 (0.17)	-1.8 (0.24)	-11	-0.05	< 0.01	< 0.01
Genital condition, %	7.5 (0.12)	8.4 (0.13)	-0.88 (0.18)	-12	-0.03	< 0.01	< 0.01
Hematological condition, %	3.9 (0.09)	3.5 (0.08)	0.40 (0.12)	10	0.02	< 0.01	< 0.01
Metabolic condition, %	6.7 (0.12)	6.6 (0.11)	0.05 (0.17)	< +/-1	0.00	0.76	< 0.01
Pregnancy, %	8.2 (0.13)	8.3 (0.13)	-0.05 (0.18)	< +/-1	0.00	0.78	< 0.01
Psychiatric condition, %	24 (0.20)	22 (0.19)	1.9 (0.28)	7.8	0.04	< 0.01	< 0.01
Renal condition, %	6.4 (0.12)	7.9 (0.12)	-1.5 (0.18)	-23	-0.06	< 0.01	< 0.01
Skeletal condition, %	20 (0.19)	21 (0.18)	-0.86 (0.27)	-4.3	-0.02	< 0.01	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Skin condition, %	13 (0.16)	14 (0.16)	-0.58 (0.23)	-4.5	-0.02	0.01	< 0.01
Substance abuse, %	19 (0.18)	14 (0.17)	4.3 (0.25)	23	0.11	< 0.01	< 0.01
Acute ACSC, %	54 (0.23)	54 (0.23)	0.00 (0.34)	< +/-1	0.00	1.00	< 0.01
Chronic ACSC, %	46 (0.23)	46 (0.23)	0.00 (0.34)	< +/-1	0.00	1.00	< 0.01
Primary diagnosis of ACSC, %	68 (0.22)	68 (0.23)	0.00 (0.31)	< +/-1	0.00	1.00	< 0.01
Only secondary diagnosis of ACSC, %	32 (0.22)	32 (0.23)	0.00 (0.31)	< +/-1	0.00	1.00	< 0.01
Primary ACSC diagnosis, %							
None	32 (0.22)	32 (0.23)	0.00 (0.31)	< +/-1	0.00	1.00	< 0.01
Angina	0.12 (0.02)	0.18 (0.02)	-0.05 (0.03)	-43	-0.02	0.04	< 0.01
Asthma	10 (0.14)	11 (0.12)	-0.80 (0.20)	-7.7	-0.03	< 0.01	< 0.01
Chest pain	20 (0.19)	16 (0.16)	3.7 (0.25)	19	0.09	< 0.01	< 0.01
Cellulitis	5.3 (0.11)	5.8 (0.11)	-0.49 (0.15)	-9.3	-0.02	< 0.01	< 0.01
COPD	0.77 (0.04)	1.1 (0.04)	-0.32 (0.06)	-41	-0.04	< 0.01	< 0.01
Dehydration	3.7 (0.09)	5.5 (0.11)	-1.8 (0.14)	-48	-0.09	< 0.01	< 0.01
Diabetes	3.9 (0.09)	3.9 (0.07)	0.02 (0.13)	< +/-1	0.00	0.90	< 0.01
DVT	0.08 (0.01)	0.07 (0.01)	0.01 (0.02)	11	0.00	0.61	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Heart failure	0.10 (0.01)	0.18 (0.02)	-0.08 (0.02)	-77	-0.02	< 0.01	< 0.01
Hypertension	3.9 (0.09)	4.7 (0.08)	-0.77 (0.14)	-20	-0.04	< 0.01	< 0.01
Intractable pain	3.7 (0.09)	2.0 (0.05)	1.8 (0.11)	47	0.09	< 0.01	< 0.01
Pneumonia	1.3 (0.05)	1.5 (0.05)	-0.19 (0.08)	-14	-0.02	0.01	< 0.01
Seizure	2.8 (0.08)	1.9 (0.06)	0.91 (0.10)	32	0.05	< 0.01	< 0.01
Sickle cell	0.43 (0.03)	0.24 (0.02)	0.19 (0.04)	44	0.03	< 0.01	< 0.01
Syncope	3.0 (0.08)	2.6 (0.07)	0.37 (0.11)	12	0.02	< 0.01	< 0.01
UTI	8.2 (0.13)	11 (0.14)	-2.5 (0.20)	-30	-0.09	< 0.01	< 0.01
Secondary ACSC diagnosis, %							
None	68 (0.22)	68 (0.23)	0.00 (0.31)	< +/-1	0.00	1.00	< 0.01
Angina	0.02 (0.01)	0.02 (0.01)	-0.01 (0.01)	-46	-0.01	0.44	< 0.01
Asthma	3.5 (0.09)	4.6 (0.12)	-1.1 (0.14)	-32	-0.06	< 0.01	< 0.01
Chest pain	3.7 (0.09)	3.2 (0.08)	0.59 (0.12)	16	0.03	< 0.01	< 0.01
Cellulitis	0.27 (0.02)	0.33 (0.02)	-0.06 (0.04)	-21	-0.01	0.14	< 0.01
COPD	0.19 (0.02)	0.23 (0.03)	-0.04 (0.03)	-22	-0.01	0.17	< 0.01
Dehydration	0.83 (0.04)	0.96 (0.05)	-0.12 (0.06)	-15	-0.01	0.04	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Diabetes	5.8 (0.11)	6.2 (0.13)	-0.48 (0.16)	-8.2	-0.02	< 0.01	< 0.01
DVT	0.06 (0.01)	0.08 (0.01)	-0.02 (0.02)	-30	-0.01	0.32	< 0.01
Heart failure	0.05 (0.01)	0.05 (0.01)	0.01 (0.01)	12	0.00	0.68	< 0.01
Hypertension	11 (0.14)	11 (0.16)	-0.36 (0.21)	-3.4	-0.01	0.08	< 0.01
Intractable pain	2.9 (0.08)	0.95 (0.06)	2.0 (0.09)	67	0.12	< 0.01	< 0.01
Pneumonia	0.14 (0.02)	0.09 (0.01)	0.05 (0.02)	36	0.01	0.03	< 0.01
Seizure	0.78 (0.04)	0.67 (0.04)	0.10 (0.06)	13	0.01	0.08	< 0.01
Sickle cell	0.10 (0.01)	0.06 (0.01)	0.04 (0.02)	40	0.01	0.04	< 0.01
Syncope	0.72 (0.04)	0.68 (0.04)	0.03 (0.06)	4.8	0.00	0.53	< 0.01
UTI	2.7 (0.08)	3.3 (0.08)	-0.57 (0.11)	-21	-0.04	< 0.01	< 0.01
Service use							
Total hospitalizations	579 (8.7)	485 (5.6)	94 (11)	16	0.05	< 0.01	< 0.01
Total hospitalizations, 3 months before enrollment	677 (12)	567 (8.3)	110 (16)	16	0.04	< 0.01	< 0.01
Total hospitalizations (top-coded)[b]	486 (5.6)	435 (4.4)	51 (7.7)	11	0.04	< 0.01	< 0.01
Total hospitalizations, 3 months before enrollment (top coded)[b]	557 (8.1)	491 (6.6)	66 (11)	12	0.04	< 0.01	< 0.01
Total ED or observation visits	3,261 (30)	2,841 (16)	420 (36)	13	0.07	< 0.01	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Total ED or observation visits, 3 months before enrollment	6,808 (42)	6,131 (23)	676 (50)	9.9	0.08	< 0.01	< 0.01
Primary care visits, any setting	3,696 (35)	3,975 (31)	-279 (47)	-7.6	-0.04	< 0.01	< 0.01
Primary care visits, any setting, day of enrollment	2.7 (0.08)	3.3 (0.08)	-0.60 (0.12)	-22	-0.04	< 0.01	< 0.01
Primary care visits, any setting, day of enrollment (top-coded)	10,681 (316)	13,265 (355)	-2,584 (484)	-24	-0.04	< 0.01	< 0.01
Primary care visits, any setting, 3 months before enrollment	4,332 (45)	4,975 (44)	-643 (66)	-15	-0.07	< 0.01	< 0.01
Primary care visits, any setting, 3 months before enrollment (top coded)[b]	3,888 (31)	4,510 (31)	-622 (46)	-16	-0.10	< 0.01	< 0.01
Primary care visits, ambulatory setting	2,600 (22)	2,729 (23)	-129 (30)	-5.0	-0.03	< 0.01	< 0.01
Primary care visits, ambulatory setting, day of enrollment	6,038 (242)	7,599 (284)	-1,561 (375)	-26	-0.03	< 0.01	< 0.01
Specialist visits, any setting	8,340 (68)	8,585 (58)	-245 (94)	-2.9	-0.02	< 0.01	< 0.01
Specialist visits, any setting, day of enrollment	25,458 (544)	17,034 (513)	8,425 (719)	33	0.07	< 0.01	< 0.01
Specialist visits, any setting, day of enrollment (top coded)	5.7 (0.11)	3.7 (0.09)	2.0 (0.14)	36	0.09	< 0.01	< 0.01
Specialist visits, any setting, 3 months before enrollment	10,464 (93)	9,928 (81)	535 (123)	5.1	0.03	< 0.01	< 0.01
Specialist visits, any setting (top-coded)[b]	7,723 (49)	8,087 (46)	-364 (70)	-4.7	-0.04	< 0.01	< 0.01
Specialist visits, any setting, day of enrollment (top-coded)[b]	20,920 (399)	13,486 (344)	7,433 (518)	36	0.09	< 0.01	< 0.01
Specialist visits, ambulatory setting	5,197 (41)	5,864 (42)	-666 (61)	-13	-0.08	< 0.01	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Specialist visits, ambulatory setting, day of enrollment	9,650 (366)	8,603 (400)	1,047 (510)	11	0.01	0.04	< 0.01
Outpatient ED and observation visits (top-coded)[b]	2,891 (14)	2,656 (11)	235 (19)	8.1	0.08	< 0.01	< 0.01
Outpatient ED and observation visits, 3 months before enrollment (top-coded)[b]	6,317 (22)	5,920 (17)	397 (29)	6.3	0.08	< 0.01	< 0.01
Total index stays	361 (6.1)	313 (4.1)	48 (8.0)	13	0.04	< 0.01	< 0.01
Total 30-day unplanned readmissions	120 (4.3)	84 (2.5)	36 (5.3)	30	0.04	< 0.01	< 0.01
AHRQ CCS category, %							
None	63 (0.23)	67 (0.23)	-3.2 (0.32)	-5.0	-0.07	< 0.01	< 0.01
Abdominal pain	2.5 (0.07)	3.5 (0.07)	-0.94 (0.11)	-37	-0.06	< 0.01	< 0.01
Headache, including migraine	1.4 (0.06)	2.4 (0.06)	-0.97 (0.09)	-69	-0.08	< 0.01	< 0.01
Other upper respiratory infections	1.4 (0.05)	1.9 (0.07)	-0.54 (0.09)	-39	-0.05	< 0.01	< 0.01
Other lower respiratory disease	1.8 (0.06)	2.3 (0.05)	-0.47 (0.10)	-25	-0.03	< 0.01	< 0.01
Other complications of pregnancy	1.9 (0.06)	0.97 (0.06)	0.96 (0.08)	50	0.07	< 0.01	< 0.01
Other nontraumatic joint disorders	1.7 (0.06)	1.0 (0.06)	0.65 (0.08)	39	0.05	< 0.01	< 0.01
Other connective tissue disease	1.6 (0.06)	1.2 (0.06)	0.45 (0.08)	28	0.04	< 0.01	< 0.01
Other injuries due to external causes	1.1 (0.05)	0.54 (0.05)	0.61 (0.06)	53	0.06	< 0.01	< 0.01
Spondylosis, intervertebral disc disorders	2.5 (0.07)	1.4 (0.08)	1.0 (0.09)	41	0.07	< 0.01	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Other	21 (0.19)	18 (0.20)	2.4 (0.26)	12	0.06	< 0.01	< 0.01
Area-level factors							
Percentage Hispanic in zip code of residence	39 (0.11)	41 (0.09)	-1.1 (0.15)	-2.8	-0.05	< 0.01	< 0.01
Percentage White in zip code of residence	24 (0.09)	26 (0.10)	-2.0 (0.12)	-8.4	-0.11	< 0.01	< 0.01
Percentage Asian in zip code of residence	9.5 (0.06)	10.0 (0.07)	-0.47 (0.08)	-4.9	-0.04	< 0.01	< 0.01
Percentage Black in zip code of residence	39 (0.13)	35 (0.12)	3.9 (0.16)	10	0.15	< 0.01	< 0.01
Percentage multiracial in zip code of residence	3.4 (0.01)	3.7 (0.01)	-0.27 (0.01)	-7.8	-0.13	< 0.01	< 0.01
Median county household income, 2014	47,196 (61)	47,202 (56)	-5.9 (86)	< +/-1	0.00	0.95	< 0.01
Median county household income, 2015	49,519 (61)	49,525 (53)	-6.0 (87)	< +/-1	0.00	0.95	< 0.01
Percentage high school graduates in zip code of residence	23 (0.03)	24 (0.03)	-1.1 (0.04)	-4.8	-0.20	< 0.01	< 0.01
Percentage of adults in county with four-year college degree	29 (0.05)	29 (0.06)	0.00 (0.08)	< +/-1	0.00	0.95	< 0.01
Persons 25+ years, 2011 to 2015	1,333,499 (1,746)	1,333,612 (1,346)	-113 (2,462)	< +/-1	0.00	0.96	< 0.01
Persons 25tears or older with 4 or more years of college, 2011 to 2015	407,353 (938)	407,439 (734)	-85 (1,348)	< +/-1	0.00	0.95	< 0.01
Percentage of population in poverty in zip code of residence	27 (0.05)	27 (0.04)	-0.27 (0.07)	-1.0	-0.02	< 0.01	< 0.01
Percentage uninsured in zip code of residence	16 (0.02)	16 (0.02)	0.22 (0.03)	1.4	0.04	< 0.01	< 0.01
Hospital occupancy rate in 2014	0.78 (0.00)	0.69 (0.00)	0.09 (0.00)	12	1.70	< 0.01	1.00

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Disproportionate-share hospital patient proportion	0.90 (0.00)	0.84 (0.00)	0.06 (0.00)	6.4	0.67	< 0.01	1.00
Number of beds in hospital providing treatment	323 (0.39)	244 (0.37)	79 (0.54)	25	0.95	< 0.01	1.00
Number of hospital beds in county	5,403 (7.5)	5,404 (8.9)	-0.17 (11)	< +/-1	0.00	0.99	< 0.01
Hospital in a health professional shortage area, %	91 (0.13)	90 (0.17)	1.3 (0.19)	1.4	0.04	< 0.01	< 0.01
Resides in a mental health professional shortage area, %	2.0 (0.00)	2.0 (0.00)	0.00 (0.00)	< +/-1	0.00	0.62	< 0.01
PCPs per 100,000 population in zip code of residence	259 (0.12)	259 (0.12)	0.50 (0.16)	< +/-1	0.02	< 0.01	< 0.01
Propensity score	0.16 (0.00)	0.14 (0.00)	0.02 (0.00)	12	0.25	< 0.01	0.20
Number of beneficiaries[c]	45,277	47,602					
Omnibus test				Chi-squared statistic 62,715.03	Degrees of freedom 95.00	P-value 0.00	

Sources: Mathematica's analysis of Medicaid claims data. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research-identifiable files or other data sources. Area-level factors from the Area Health Resource File 2016–2017 release.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table were calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid. Exact matching variables include quarter of enrollment, county of residence, and ACSC diagnosis.

[a] The CDPS score relied on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

Table B.2 (*continued*)

[c] Treatment-eligible Medicaid beneficiaries from the whole implementation period were used for matching. Therefore, the number of beneficiaries in this table differs from the number of beneficiaries in Tables 2 and 3 in the main body of this report, which reflect the group of beneficiaries seen at EDs within the first nine months of the program.

ACSC = ambulatory care-sensitive condition; AHRQ = Agency for Healthcare Research and Quality; CCS = clinical classification software; CDPS = Chronic Illness and Disability Payment System; COPD = chronic obstructive pulmonary disease; DVT = deep vein thrombosis; ED = emergency department, HCBS = home and community-based services; NYC H+H = New York City Health + Hospitals; PCP = primary care physician; SE = standard error; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System; UTI = urinary tract infection.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Tables C.1 through C.4 display the results from the impact analysis. Table C.1 shows the impact estimates for the full Medicare study population. Table C.2 shows similar Medicare results for the subgroup of treatment-eligible beneficiaries who enrolled within the first nine months of the program start date versus the treatment-eligible beneficiaries who enrolled after the first nine months of the launch date. Table C.3 shows the impact estimates for the full Medicaid study population. Table C.4 shows similar Medicaid results for the subgroup of treatment-eligible beneficiaries who enrolled within the first nine months of the program start date versus the treatment-eligible beneficiaries who enrolled after the first nine months of the launch date. The study estimated the models over Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that are statistically different from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the NYC H+H intervention on select Medicare FFS expenditures (dollars PBPM) and use measures during a 12-month follow-up period

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)					
Baseline year	2,160	2,003			
Months 1 to 6	3,128	2,945	26 (100)	< 1%	0.79
Months 7 to 12	2,714	2,626	-70 (95)	-2.7%	0.46
Months 1 to 12	2,945	2,800	-13 (83)	< 1%	0.88
Acute inpatient expenditures ($ PBPM)					
Baseline year	1,060	870			
Months 1 to 6	1,664	1,489	-16 (77)	< 1%	0.84
Months 7 to 12	1,409	1,300	-81 (71)	-6.1%	0.25
Months 1 to 12	1,551	1,403	-43 (63)	-2.8%	0.50
SNF expenditures ($ PBPM)					
Baseline year	178	211			
Months 1 to 6	303	304	32 (24)	12%	0.19
Months 7 to 12	300	271	62** (27)	29%	0.02
Months 1 to 12	299	288	45** (21)	18%	0.03
Professional Part B expenditures ($ PBPM)					
Baseline year	335	377			
Months 1 to 6	413	494	-39**** (10)	-8.6%	< 0.01
Months 7 to 12	387	462	-33*** (10)	-8.0%	< 0.01
Months 1 to 12	401	479	-36**** (9.0)	-8.4%	< 0.01

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Primary care visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	5,211	5,372			
Months 1 to 6	6,308	6,551	-82 (129)	-1.3%	0.53
Months 7 to 12	6,030	6,435	-245* (148)	-4.0%	0.10
Months 1 to 12	6,191	6,503	-151 (117)	-2.4%	0.20
Specialist visits in all settings, per 1,000 beneficiaries					
Baseline year	15,013	14,724			
Months 1 to 6	18,129	18,586	-746* (383)	-4.0%	0.05
Months 7 to 12	16,551	17,136	-874** (382)	-5.1%	0.02
Months 1 to 12	17,434	17,938	-793** (335)	-4.4%	0.02
Hospital stays, per 1,000 beneficiaries					
Baseline year	766	662			
Months 1 to 6	1,068	949	15 (37)	1.4%	0.69
Months 7 to 12	893	844	-55 (37)	-6.2%	0.13
Months 1 to 12	990	900	-14 (31)	-1.5%	0.65
Hospital stays for an ACSC, per 1,000 beneficiaries					
Baseline year	287	235			
Months 1 to 6	363	315	-3.6 (21)	< 1%	0.86
Months 7 to 12	290	268	-30 (20)	-9.6%	0.14
Months 1 to 12	330	293	-15 (17)	-4.3%	0.39
ED or observation visits, per 1,000 beneficiaries (top-coded)[b]					
Baseline year	2,411	2,164			
Months 1 to 6	1,846	1,643	-44 (47)	-2.3%	0.36
Months 7 to 12	1,504	1,335	-78 (50)	-5.0%	0.12
Months 1 to 12	1,707	1,515	-55 (42)	-3.1%	0.18
ED or observation visits for an ACSC, per 1,000 beneficiaries[b]					
Baseline year	666	643			
Months 1 to 6	378	356	-1.2 (19)	< 1%	0.95
Months 7 to 12	300	301	-24 (20)	-7.7%	0.23
Months 1 to 12	340	331	-14 (18)	-3.8%	0.43
Percentage of beneficiaries with any ED or observation visits in a time period[c]					
Baseline year	37	33			
Months 1 to 6	40	43	-3.3 (2.0)	-7.6%	0.10
Months 7 to 12	35	37	-2.4 (2.0)	-6.5%	0.24
Months 1 to 12	55	60	-5.3** (2.4)	-8.8%	0.03
Percentage of beneficiaries with more than two ED or observation visits in a time period[c]					
Baseline year	37	33			
Months 1 to 6	20	22	-2.5 (1.7)	-11%	0.14

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Months 7 to 12	17	20	-3.2* (1.7)	-16%	0.06
Months 1 to 12	28	34	-5.8** (2.3)	-17%	0.01
Number of index discharges for readmissions					
Baseline year	5,603	4,183			
Months 1 to 6	4,104	3,751			
Months 7 to 12	2,753	2,469			
Months 1 to 12	6,857	6,220			
Sample sizes					
Number of beneficiaries					
Baseline year	9,134	9,901			
Months 1 to 6	9,134	9,901			
Months 7 to 12	7,492	8,380			
Months 1 to 12	9,134	9,901			

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicare claims and enrollment data as of August 10, 2018.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline year and follow-up year.

[c] Because an ED visit triggered eligibility for the intervention, the proportion of treatment and matched comparison beneficiaries with any ED visit at baseline was 100 percent. To examine the impact on this outcome, the model includes controls for the probability of having more than two ED visits at baseline, instead of controlling for the baseline probability of any ED visit.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; NYC H+H = New York City Health + Hospitals; PBPM = per beneficiary per month; SE = standard error; SNF = skilled nursing facility.

Table C.2. Estimated impact of the NYC H+H intervention on select Medicare FFS expenditures (dollars PBPM) and use measures during a 12-month follow-up period by program maturity

	Beneficiaries enrolled within nine months of program start date					Beneficiaries enrolled after nine months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)										
Baseline year	2,210	2,092				2,137	1,960			
Months 1 to 6	3,152	2,889	145 (163)	4.9%	0.37	3,117	2,970	-30 (125)	< 1%	0.81
Months 7 to 12	2,743	2,758	-132 (178)	-4.9%	0.46	2,700	2,556	-33 (110)	-1.3%	0.76
Months 1 to 12	2,957	2,821	19 (145)	< 1%	0.90	2,939	2,791	-29 (101)	-1.0%	0.78
Acute inpatient expenditures ($ PBPM)										
Baseline year	1,141	956				1,021	828			
Months 1 to 6	1,753	1,484	84 (129)	5.1%	0.52	1,621	1,491	-62 (95)	-3.8%	0.51
Months 7 to 12	1,453	1,389	-121 (134)	-8.5%	0.37	1,391	1,256	-57 (82)	-4.5%	0.49
Months 1 to 12	1,611	1,436	-9.5 (113)	< 1%	0.93	1,522	1,389	-59 (75)	-4.0%	0.43
SNF expenditures ($ PBPM)										
Baseline year	165	177				184	228			
Months 1 to 6	289	256	45 (38)	19%	0.23	309	327	25 (31)	8.9%	0.41
Months 7 to 12	279	271	20 (49)	8.4%	0.69	311	269	85*** (31)	42%	< 0.01
Months 1 to 12	282	261	33 (35)	14%	0.34	307	301	50** (25)	20%	0.05
Professional Part B expenditures ($ PBPM)										
Baseline year	330	373				337	379			
Months 1 to 6	416	494	-36* (19)	-7.9%	0.06	412	495	-40**** (12)	-8.9%	< 0.01
Months 7 to 12	381	471	-48*** (18)	-12%	< 0.01	390	456	-25* (13)	-6.1%	0.06
Months 1 to 12	399	483	-41** (16)	-9.5%	0.01	402	478	-34*** (11)	-7.8%	< 0.01
Primary care visits in ambulatory settings, per 1,000 beneficiaries										
Baseline year	5,247	5,196				5,194	5,456			
Months 1 to 6	6,372	6,335	-14 (202)	< 1%	0.94	6,278	6,653	-114 (164)	-1.8%	0.49
Months 7 to 12	6,179	6,454	-326 (258)	-5.2%	0.21	5,951	6,412	-199 (181)	-3.4%	0.27
Months 1 to 12	6,285	6,392	-158 (196)	-2.5%	0.42	6,146	6,555	-148 (145)	-2.4%	0.31

Table C.2 (*continued*)

	Beneficiaries enrolled within nine months of program start date					Beneficiaries enrolled after nine months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Specialist visits in all settings, per 1,000 beneficiaries										
Baseline year	15,669	15,127				14,701	14,532			
Months 1 to 6	18,809	18,550	-283 (746)	-1.5%	0.70	17,806	18,601	-963** (442)	-5.2%	0.03
Months 7 to 12	16,978	17,141	-704 (699)	-4.1%	0.31	16,365	17,150	-954** (454)	-5.5%	0.04
Months 1 to 12	17,946	17,874	-470 (654)	-2.6%	0.47	17,192	17,975	-952** (384)	-5.3%	0.01
Hospital stays, per 1,000 beneficiaries										
Baseline year	897	709				704	639			
Months 1 to 6	1,198	968	42 (67)	3.7%	0.53	1,006	939	2.5 (45)	< 1%	0.95
Months 7 to 12	996	894	-86 (68)	-8.6%	0.21	846	819	-37 (43)	-4.5%	0.38
Months 1 to 12	1,101	930	-17 (59)	-1.5%	0.78	937	886	-13 (37)	-1.4%	0.72
Hospital stays for an ACSC, per 1,000 beneficiaries										
Baseline year	354	277				256	216			
Months 1 to 6	441	328	37 (36)	9.3%	0.30	326	308	-22 (25)	-6.4%	0.37
Months 7 to 12	361	293	-8.4 (36)	-2.4%	0.82	256	257	-41* (24)	-14%	0.09
Months 1 to 12	403	310	16 (31)	4.3%	0.60	296	285	-30 (20)	-9.1%	0.14
ED or observation visits, per 1,000 beneficiaries (top-coded)[b]										
Baseline year	2,792	2,636				2,230	1,939			
Months 1 to 6	2,115	1,980	-21 (92)	-1.0%	0.82	1,718	1,481	-54 (55)	-3.1%	0.32
Months 7 to 12	1,675	1,615	-97 (99)	-5.6%	0.33	1,436	1,211	-66 (55)	-4.5%	0.23
Months 1 to 12	1,928	1,806	-35 (83)	-1.8%	0.68	1,604	1,378	-65 (47)	-3.9%	0.16
ED or observation visits for an ACSC, per 1,000 beneficiaries[b]										
Baseline year	783	799				610	568			
Months 1 to 6	447	475	-12 (38)	-2.7%	0.75	345	299	3.8 (21)	1.1%	0.86
Months 7 to 12	355	390	-19 (40)	-5.3%	0.64	277	262	-28 (23)	-9.3%	0.23
Months 1 to 12	403	433	-14 (36)	-3.2%	0.70	311	283	-14 (19)	-4.2%	0.47

Table C.2 (*continued*)

	Beneficiaries enrolled within nine months of program start date					Beneficiaries enrolled after nine months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Percentage of beneficiaries with any ED or observation visits in a time period[c]										
Baseline year	44	42				34	28			
Months 1 to 6	43	48	-4.3* (2.3)	-9.2%	0.06	39	42	-2.7 (2.1)	-6.6%	0.20
Months 7 to 12	36	39	-2.8 (2.4)	-7.2%	0.23	33	36	-2.1 (2.2)	-5.9%	0.33
Months 1 to 12	58	63	-5.7** (2.6)	-9.1%	0.03	54	59	-5.0** (2.5)	-8.5%	0.04
Percentage of beneficiaries with more than two ED or observation visits in a time period[c]										
Baseline year	44	42				34	28			
Months 1 to 6	21	25	-3.5* (1.9)	-14%	0.07	19	21	-1.9 (1.7)	-9.0%	0.28
Months 7 to 12	17	21	-4.6** (1.9)	-22%	0.02	17	19	-2.2 (1.7)	-12%	0.21
Months 1 to 12	29	36	-7.6*** (2.6)	-21%	< 0.01	28	33	-4.7** (2.4)	-14%	0.05
Number of index discharges for readmissions										
Baseline year	2,127	1,492				3,476	2,691			
Months 1 to 6	1,447	1,181				2,657	2,570			
Months 7 to 12	1,003	840				1,750	1,629			
Months 1 to 12	2,450	2,021				4,407	4,199			
Sample sizes										
Number of beneficiaries										
Baseline year	2,925	3,153				6,209	6,748			
Months 1 to 6	2,925	3,153				6,209	6,748			
Months 7 to 12	2,588	2,830				4,904	5,550			
Months 1 to 12	2,925	3,153				6,209	6,748			

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicare claims and enrollment data as of August 10, 2018.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline year and follow-up year.

Table C.2 (*continued*)

[c] Because an ED visit triggered eligibility for the intervention, the proportion of treatment and matched comparison beneficiaries with any ED visit at baseline was 100 percent. For examining impact on this outcome, the model includes controls for the probability of having more than two ED visits at baseline, instead of controlling for the baseline probability of any ED visit.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; NYC H+H = New York City Health + Hospitals; PBPM = per beneficiary per month, SNF = skilled nursing facility; SE = standard error.

Table C.3. Estimated impact of the NYC H+H intervention on select Medicaid use measures during a 12-month follow-up period

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Primary care visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	2,659	2,782			
Months 1 to 6	3,162	3,434	-149*** (47)	-4.6%	< 0.01
Months 7 to 12	2,777	2,994	-94* (53)	-3.3%	0.07
Months 1 to 12	2,996	3,245	-126*** (43)	-4.1%	< 0.01
Specialist visits in all settings, per 1,000 beneficiaries					
Baseline year	8,426	8,654			
Months 1 to 6	9,987	10,249	-34 (156)	< 1%	0.83
Months 7 to 12	8,594	8,929	-107 (164)	-1.2%	0.51
Months 1 to 12	9,384	9,680	-68 (137)	< 1%	0.62
Hospital stays, per 1,000 beneficiaries[b]					
Baseline year	576	481			
Months 1 to 6	758	612	51*** (15)	7.3%	< 0.01
Months 7 to 12	596	516	-15 (17)	-2.5%	0.38
Months 1 to 12	688	571	22 (14)	3.3%	0.11
Hospital stays for an ACSC, per 1,000 beneficiaries[b]					
Baseline year	157	132			
Months 1 to 6	202	174	1.9 (7.7)	< 1%	0.81
Months 7 to 12	154	146	-18** (8.9)	-11%	0.04
Months 1 to 12	181	162	-6.8 (6.9)	-3.7%	0.32
ED or observation visits, per 1,000 beneficiaries (top-coded)[b]					
Baseline year	2,859	2,605			
Months 1 to 6	2,387	2,096	37 (29)	1.6%	0.20
Months 7 to 12	1,958	1,800	-96*** (32)	-4.7%	< 0.01
Months 1 to 12	2,202	1,975	-27 (26)	-1.2%	0.31
ED or observation visits for an ACSC, per 1,000 beneficiaries[b]					
Baseline year	1,098	1,028			
Months 1 to 6	595	519	6.3 (12)	1.1%	0.61
Months 7 to 12	461	427	-36*** (13)	-7.3%	< 0.01
Months 1 to 12	534	479	-15 (11)	-2.7%	0.17
Percentage of beneficiaries with any ED or observation visits in a time period[c]					
Baseline year	46	44			
Months 1 to 6	47	49	-1.4 (0.92)	-3.0%	0.12
Months 7 to 12	40	43	-3.0*** (0.94)	-7.0%	< 0.01
Months 1 \ to 12	62	64	-1.5 (1.1)	-2.4%	0.15

Table C.3 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Percentage of beneficiaries with more than two ED or observation visits in a time period[c]					
Baseline year	46	44			
Months 1 to 6	25	26	-0.68 (0.80)	-2.6%	0.39
Months 7 to 12	22	24	-2.2*** (0.81)	-8.9%	< 0.01
Months 1 to 12	36	37	-1.0 (1.1)	-2.8%	0.33
Sample sizes					
Number of beneficiaries					
Baseline year	45,277	47,602			
Months 1 to 6	45,277	47,602			
Months 7 to 12	38,555	40,789			
Months 1 to 12	45,277	47,602			

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicaid claims and enrollment data as of August 10, 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research-identifiable files or other data sources.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline year and follow-up year.

[c] Because an ED visit triggered eligibility for the intervention, the proportion of treatment and matched comparison beneficiaries with any ED visit at baseline was 100 percent. To examine the impact on this outcome, the model includes controls for the probability of having more than two ED visits at baseline, instead of controlling for the baseline probability of any ED visit.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ACSC = ambulatory care-sensitive condition; ED = emergency department; NYC H+H = New York City Health + Hospitals; SE = standard error; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System.

Table C.4. Estimated impact of the NYC H+H intervention on select Medicaid use measures during a 12-month follow-up period by program maturity

	Beneficiaries enrolled within nine months of program start date					Beneficiaries enrolled after nine months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Primary care visits in ambulatory settings, per 1,000 beneficiaries										
Baseline year	3,163	3,161				2,530	2,685			
Months 1 to 6	3,155	3,480	-326*** (104)	-9.6%	< 0.01	3,163	3,422	-104* (54)	-3.2%	0.05
Months 7 to 12	2,745	2,971	-228* (126)	-8.0%	0.07	2,797	3,010	-58 (57)	-2.0%	0.31
Months 1 to 12	2,963	3,242	-281*** (98)	-8.9%	< 0.01	3,009	3,249	-85* (47)	-2.8%	0.07
Specialist visits in all settings, per 1,000 beneficiaries										
Baseline year	12,153	11,527				7,472	7,917			
Months 1 to 6	11,495	11,423	-555 (372)	-4.7%	0.14	9,597	9,943	98 (172)	1.0%	0.57
Months 7 to 12	9,352	9,311	-586 (360)	-6.3%	0.10	8,473	8,899	19 (184)	< 1%	0.92
Months 1 to 12	10,506	10,446	-567* (317)	-5.4%	0.07	9,119	9,502	62 (152)	< 1%	0.68
Hospital stays, per 1,000 beneficiaries										
Baseline year	918	743				488	413			
Months 1 to 6	986	812	-2.0 (38)	< 1%	0.96	700	561	64*** (17)	10%	< 0.01
Months 7 to 12	779	726	-122*** (43)	-14%	< 0.01	553	463	15 (19)	2.8%	0.42
Months 1 to 12	890	773	-57* (34)	-6.3%	0.09	638	520	43*** (15)	7.3%	< 0.01
Hospital stays for an ACSC, per 1,000 beneficiaries										
Baseline year	284	247				125	102			
Months 1 to 6	322	284	0.95 (21)	< 1%	0.96	171	146	2.0 (8.2)	1.2%	0.81
Months 7 to 12	239	237	-35 (22)	-14%	0.11	134	124	-13 (9.5)	-8.9%	0.17
Months 1 to 12	283	262	-16 (18)	-5.6%	0.39	155	137	-4.5 (7.3)	-2.8%	0.54
ED or observation visits, per 1,000 beneficiaries (top-coded)[b]										
Baseline year	3,471	3,182				2,702	2,457			
Months 1 to 6	2,486	2,363	-166** (69)	-6.4%	0.02	2,362	2,028	89*** (32)	4.0%	< 0.01
Months 7 to 12	2,197	2,160	-253*** (77)	-10%	< 0.01	1,902	1,711	-54 (35)	-2.8%	0.13
Months 1 to 12	2,369	2,293	-213**** (62)	-8.2%	< 0.01	2,163	1,896	22 (29)	1.0%	0.44

Table C.4 (*continued*)

	Beneficiaries enrolled within nine months of program start date					Beneficiaries enrolled after nine months of program start date				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
ED or observation visits for an ACSC, per 1,000 beneficiaries[b]										
Baseline year	1,356	1,264				1,032	967			
Months 1 to 6	681	630	-41 (29)	-5.9%	0.16	573	490	18 (14)	3.4%	0.18
Months 7 to 12	568	569	-92*** (30)	-14%	< 0.01	436	392	-21 (15)	-4.5%	0.16
Months 1 to 12	623	606	-74*** (26)	-10%	< 0.01	513	448	0.05 (12)	< 1%	1.00
Percentage of beneficiaries with any ED or observation visits in a time period[c]										
Baseline year	53	51				45	42			
Months 1 to 6	47	50	-3.3*** (1.3)	-6.6%	< 0.01	47	48	-1.1 (0.95)	-2.2%	0.26
Months 7 to 12	42	47	-4.5*** (1.3)	-9.7%	< 0.01	40	43	-2.7*** (0.97)	-6.4%	< 0.01
Months 1 to 12	62	67	-4.3*** (1.4)	-6.5%	< 0.01	62	63	-1.0 (1.1)	-1.6%	0.35
Percentage of beneficiaries with more than two ED or observation visits in a time period[c]										
Baseline year	53	51				45	42			
Months 1 to 6	26	28	-1.8 (1.1)	-6.5%	0.11	25	25	-0.46 (0.82)	-1.8%	0.58
Months 7 to 12	23	26	-2.4** (1.1)	-9.3%	0.04	22	24	-2.2*** (0.84)	-9.0%	< 0.01
Months 1 to 12	36	39	-2.1 (1.4)	-5.6%	0.12	36	37	-0.85 (1.1)	-2.3%	0.43
Sample sizes										
Number of beneficiaries										
Baseline year	9,747	9,995				35,530	37,607			
Months 1 to 6	9,747	9,995				35,530	37,607			
Months 7 to 12	8,727	9,173				29,828	31,616			
Months 1 to 12	9,747	9,995				35,530	37,607			

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicaid claims and enrollment data as of August 10, 2018. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research-identifiable files or other data sources.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline year and follow-up year.

[c] Because an ED visit triggered eligibility for the intervention, the proportion of treatment and matched comparison beneficiaries with any ED visit at baseline was 100 percent. To examine impact on this outcome, the model includes controls for the probability of having more than two ED visits at baseline, instead of controlling for the baseline probability of any ED visit.

Table C.4 (*continued*)

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.
ACSC = ambulatory care-sensitive condition; ED = emergency department; NYC H+H = New York City Health + Hospitals; SE = standard error; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System.

Appendix D

Results from Bayesian Analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the study estimated program impacts for New York City Health + Hospitals (NYC H+H) using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which makes decision making easier by summarizing both the size and the certainty of an impact in a single value. Drawing probabilistic conclusions requires external or prior evidence. In this analysis, the findings from the evaluation of 87 awardees included in the Round 1 of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to NYC H+H. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on three core outcomes for Medicare, and two core outcome for Medicaid and the Medicare and Medicaid pooled estimates, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report, separately for Medicare and Medicaid populations. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for NYC H+H led to a Bayesian estimate of the program's impact on total Medicare expenditures of less than -1 percent (an estimated reduction of $6 per beneficiary per month) in the first year; this calculation includes information from the Medicare population only.

Table D.1. Comparison of frequentist and Bayesian impact estimates for NYC H+H in the first year after enrollment

		Impact estimate (95 percent interval)		Percentage impacts		
Payer	Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Medicaid	Hospital admissions	22 (-4.7, 49)	9.8 (-40, 58)	1%	3%	1%
	ED visits	-27 (-78, 25)	22 (-143, 187)	< 1%	-1%	< 1%
Medicare	Total expenditures ($ PBPM)	-13 (-176, 150)	-5.8 (-151, 141)	1%	>-1%	>-1%
	Hospital admissions	-14 (-76, 47)	0.36 (-51, 52)	2%	-1%	< 1%
	ED visits	-55 (-137, 26)	-8.1 (-96, 82)	1%	-3%	>-1%

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicare and Medicaid claims and enrollment data as of August 10, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending and are evaluated in the Medicare sample only. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

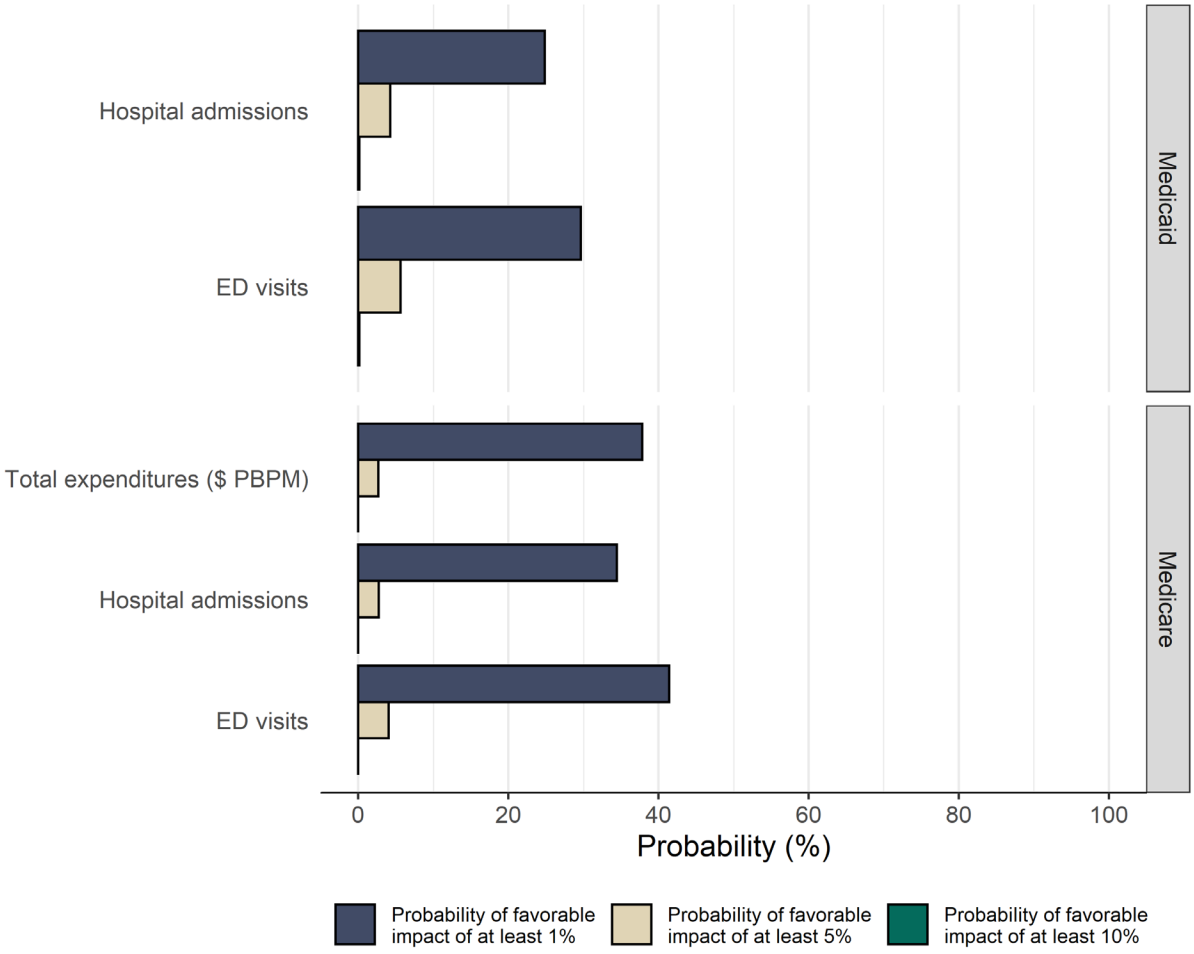
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results are somewhat imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates that are more consistent across outcomes and populations. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that most impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that NYC H+H achieved favorable impacts during the first year on three core outcomes at three thresholds: a favorable impact of 1, 5, or 10 percent or more.

Figure D.1. Probability that the NYC H+H program had a favorable impact on key outcomes



Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicare and Medicaid claims and enrollment data as of August 10, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending and are evaluated in the Medicare sample only. The Bayesian regression also

Figure D.1 (*continued*)

incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a small probability—in the range of 20 to 30 percent—that NYC H+H had a favorable impact of 1 percent or more on hospital admissions and emergency department visits in the Medicaid sample, with slightly higher probabilities, in the range of 30 to 40 percent, in the Medicare sample. These probabilities are not large enough to indicate a substantial impact. Therefore, the Bayesian analysis corroborates the findings from the frequentist analysis that the NYC H+H program did not have a meaningful impact on total Medicare expenditures or service use.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Northwell Health

September 2020

Lisa M. Lines, Arnold Chen, Justin Kirschner, Bob Schmitz, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

NORTHWELL HEALTH

Northwell Health, an integrated health system in New York, formerly known as North Shore Long Island Jewish Health System, used its Round 2 Health Care Innovation Award to expand the pilot of its Healthy Transitions in Late Stage Chronic Kidney Disease (CKD) program. The program sought to assist residents of the New York City boroughs of Manhattan and Queens and Nassau and Suffolk Counties with Stage 4 or 5 CKD. The Healthy Transitions model shifted care from a conventional nephrologist care model to greater reliance on registered nurse (RN) care managers. It included home visits by nurse care managers to determine whether patients had enough family support and an adequate home environment to manage their care. Nurse care managers also educated patients about diet and exercise to slow the progress of kidney disease and delay the start of dialysis. They also promoted transplantation, which is the preferred outcome for those who do not choose conservative care. The program launched in November 2014, three months after award, and ended in November 2017. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program aimed to reduce unnecessary hospital and ED use and lower costs for beneficiaries in four boroughs and counties in New York State with Stage 4 or 5 CKD through disease management activities.
- The program represented an expansion of an existing program that two nephrology practices in the Northwell Health network had implemented.
- Due to small sample size and underreporting of CKD in claims data, it was not possible to conduct an impact evaluation of the Health Transitions program.
- This report describes the baseline demographic and health characteristics of the 203 participants enrolled in fee-for-service (FFS) Medicare, representing 29 percent of the 705 total participants.

The awardee hypothesized that intervening earlier in the progression of kidney disease than is typical with conventional care would enable patients to choose optimal modes of renal replacement therapy, thereby improving their quality of life, reducing hospital and emergency department (ED) use, and lowering health care costs. The earlier intervention would help to avoid emergency initiation of hemodialysis through hemodialysis catheters in crisis situations. Dialysis catheters have a high risk of infection and other medical problems. The awardee expected the program to lead to more arteriovenous (AV) fistula placements,[1] preemptive kidney transplants, and home dialysis, leading to reduced use of inpatient and ED services and 30 percent lower Medicare costs.

Northwell worked with the National Kidney Foundation and the Renal Physicians Association to develop a payment model. The Association, in turn, approached the Physician-Focused Payment Model Technical Advisory Committee with a payment model for late-stage kidney disease

[1] AV fistulae enable clinicians to aspirate blood in veins, which are easier to reach than arteries. AV fistulae have many benefits over dialysis catheters, including reduced infection, clotting, and other complications.

similar to the one that Northwell helped develop that would provide the appropriate incentives for programs like Healthy Transitions. Northwell program leaders were exploring funding from a private company at the end of the award, but were not hopeful about the prospects for developing a payment model that would support a wider implementation of the Healthy Transitions program.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Healthy Transitions was a patient-centered program that aimed to integrate and coordinate all aspects of care for people with late-stage CKD by (1) focusing on patient education and care management to delay the onset of ESRD and (2) helping patients make informed choices about ESRD treatment that reflect their personal preferences.
Major innovation	The main innovation in Healthy Transitions was intervening early enough in the disease process to enable patients to make better informed and appropriate decisions about their CKD treatment options.
Program components	• Care management by RN care managers • Shared decision making • Daily weight monitoring • Quarterly home visits to discuss dialysis and transplantation options; review medications and nutrition; screen for depression, anxiety, and quality of life; and provide care management
Target population	Eligible participants were individuals with late-stage CKD (Stages 4 and 5) who (1) were at least 18 years old, (2) lived in one of four counties or boroughs in the New York City area, (3) had an estimated glomerular filtration rate of less than 30 mL/min, and (4) had no clinically apparent cognitive impairment.
Total enrollment	The awardee enrolled 705 participants, representing 141 percent of the original enrollment goal.
Level of engagement	Interviews with program staff and self-reported awardee data indicated that the program successfully engaged participants. Staff rated patient engagement as excellent and, based on an internal survey, more than 95 percent of participants said they were satisfied with the program.
Theory of change or theory of action	The awardee focused on changing participants' and providers' behavior by shifting the nephrologist-based care model to a greater reliance on nurse care managers, who can develop more personal relationships with the participants and guide them through the complex care system. The awardee hypothesized that this improved model of disease management would better prepare patients for ESRD and choose optimal modes of renal replacement therapy, which would lead to better outcomes such as improved quality of life for patients, decreased hospital and ED use, and lower health care costs.
Award amount	$2,453,742
Effective launch date	November 17, 2014
Program settings	Patients' homes; nephrology clinics and practice offices
Market area	Urban and suburban: The New York City boroughs of Manhattan and Queens and Nassau and Suffolk Counties

Table 1 (*continued*)

Program characteristics	Description
Target outcomes	• Patients better prepared for ESRD care • Increase in dialysis modality selection, home or peritoneal dialysis or preemptive transplantation, fewer in-hospital dialysis starts • Improved quality of life (KDQOL scores) • Reductions in hospitalizations and ED visits • Savings to Medicare
Payment model	Northwell collaborated with the National Kidney Foundation to develop a payment model that combined a condition-specific, population-based payment with value-based incentives and penalties for nephrologists.
Sustainability plans	After its award ended in February 2018, all but one of Northwell's sites continued the Healthy Transitions program with internal funding.

CKD = chronic kidney disease; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; KDQOL = Kidney Disease Quality of Life; RN = registered nurse.

It was not possible to conduct a rigorous impact evaluation of the Healthy Transitions program because of the way in which the awardee identified and recruited participants into the program. As a result, this report describes only the demographic and health characteristics of Medicare FFS participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis. Appendix A, Table A.1 describes the identification of the sample used for the descriptive analysis.

Table 2. Key features of the descriptive analysis

Evaluation features	Description
Descriptive analysis	Due to small sample size and lack of clinical measures socioeconomic factors used to determine eligibility in claims data, a rigorous impact evaluation of this program was not possible. As a result, this report describes only the baseline demographic and health characteristics of participants enrolled in FFS Medicare.
Intervention group used for descriptive analysis	The intervention group for the descriptive analysis relied on the 203 participants (among the total enrollment of 705) enrolled in Medicare FFS for at least three months before and after enrollment into the program, representing 29 percent of total enrollment. The 502 excluded beneficiaries included 355 patients with missing enrollment data, 90 patients who were not enrolled in Medicare or could not be identified in Medicare enrollment files, and 57 Medicare beneficiaries who were not enrolled in Parts A and B for at least three months before and after enrollment.
Limitations	Due to the problems noted above, this report cannot make inferences about the impact of this program on Medicare costs or other program outcomes.

FFS = fee-for-service.

PROGRAM DESIGN AND ADAPTATION

The Northwell Health Transitions program had two key components: management and shared decision making.[2]

Care management

Nurse care managers identified potential program participants from patients' charts, conferred with nephrologists about the patients' eligibility, and determined whether patients were appropriate for the program. They conducted an initial home visits within two weeks of enrollment to assess participants' home environments, social support needs, and risk of readmission with a risk-stratification score instrument developed for the Healthy Transitions program. During the home visit, the nurse care managers also educated patients about diet, exercise, and medication adherence to slow the progress of the disease and delay the start of dialysis for as long as possible. After this initial visit, nurses regularly followed up with participants, primarily by phone. The calls included medication reconciliation, monitoring and tracking of disease progression, continued education about the disease, and referrals to social supports or to the program's social worker.

Participants also received scales and instructions to weigh themselves and report their weight daily using an automated telephone system. The awardee incorporated these data into daily tracking reports for the nurse care managers that included participants' weight trends, upcoming and previous scheduled visits, most recent eGFR[3] level, risk stratification, and any notes or comments. Nurse care managers followed up when a participant's weight suddenly increased, which could indicate further disease progression or kidney failure.

Shared decision making

Nurse care managers also educated participants about different renal replacement therapy (RRT) modalities. These included primarily hemodialysis in a center, peritoneal dialysis at home, and kidney transplantation. Shortly after implementation, the program also began educating patients about the conservative care option: foregoing dialysis or transplantation and seeking palliative care only. If a participant's eGFR level dropped below 20 ml/min, the care manager began to work with the participant to choose an RRT modality and prepare for the modality chosen. The patient education and shared decision making before patients' kidneys reached end-stage would enable patients to make better choices about their treatment options. Having patients involved in planning early in the process allowed extra time for patients who chose hemodialysis to schedule surgery to create an AV fistula to prepare for hemodialysis.

---

[2] The Third Annual Evaluation Report provides additional details on the design and implementation of the Healthy Transitions program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

[3] The eGFR is a blood test that indicates how well the kidneys are filtering. An eGFR of 60 ml/min or higher is in the normal range, below 60 ml/min signifies kidney disease, and 15 ml/min or lower means kidney failure.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee successfully achieved its enrollment goal, exceeding its projected number of participants in Healthy Transitions during the award period. The awardee also succeeded in delivering services, recruiting and engaging providers, and engaging participants, and was mostly successful in hiring and training staff. Several factors contributed to these successes. First, Healthy Transitions was free to participants, mitigating one of the main concerns—costs—raised by patients during recruitment. Second, the program actively engaged referring nephrologists. Nurse care managers sent reminders to referring nephrologists about patients who were potentially eligible and updates on the condition of patients enrolled in the program. Starting in the second program year, nurse care managers were embedded in nephrologists' offices, where they supported the nephrologists and participated in consultations with patients about the program. Third, nurse care managers spent an extensive amount of time communicating the benefits of the program to potential enrollees and communicating with patients after enrollment. Their dedication gained the trust of patients and participating nephrologists. Lastly, the awardee continued to seek ways to support patients and providers. During the third program year, program staff began to meet with staff from the Northwell Health transplant center to encourage a team-based approach to patients' care and preemptive transplantation, and they tracked Health Transitions patients through the transplant process. The awardee also formed a medical advisory board to help program leaders and nephrologists address challenges for their patients, including how to help patients better navigate the health system.

Despite these successes, the awardee faced several implementation challenges. Staff reported having difficulty recruiting and engaging patients. The awardee estimated that about half of all eligible patients initially declined enrollment. Patients noted socioeconomic barriers and lack of supportive home environments as important factors in their decision. Some patients hesitated to enroll out of embarrassment for their home or its condition, given the required home visits. These factors also influenced service delivery and might have adversely affected program effectiveness. For example, some participants' homes were not suitable for home dialysis because of lack of space, difficulty delivering supplies, and lack of adequate cleanliness, which reduced their treatment options. Other patient engagement challenges included language barriers, low patient health literacy, lack of health education, and not feeling ill at the time of recruitment.

Other factors likely affected service delivery and program effectiveness. A large majority (90 percent) of people with CKD are undiagnosed. Many patients referred to the program were on the brink of needing dialysis, and thus much farther in the disease process than the target population assumed in the theory of action. Practices referred these patients, and the program accepted them, because the awardee believed that nurse care managers' services could still benefit these patients. However, they were too late in the disease process to fully benefit from the program's preventive and planning aspects. Another factor affecting service delivery and program effectiveness was that a major peritoneal dialysis fluid manufacturer ceased operations,

causing the peritoneal dialysis program to stop. Further concerns included health information technology barriers and difficulty working across collaborating organizations.

Finally, the awardee experienced staffing challenges as well. The nurse care manager assigned to the two Manhattan practices left her job, which ultimately prevented implementing the program at those practices. These practices had few patients, so her departure did not severely affect enrollment, but it limited the program's geographic reach. The palliative care nephrologist left in the third program year, which resulted in the awardee suspending the conservative care option. Some nurse care managers also expressed concern in interviews that their caseloads were larger (85 to 107 patients) than they believed was optimal (about 75 patients). They felt that the quality of their services would deteriorate as their caseloads grew and would ultimately become unmanageable at the high end of the range.

Implications of program implementation for achieving program goals

- Many patients were referred into the program too late in their disease progression when they were on the brink of needing dialysis to fully benefit from the intervention services.
- Because the palliative care nephrologist left the program early, the awardee had to suspend the conservative care option in the third program year.
- A number of external factors, such as low patient health literacy, language barriers, poor condition of patients' housing, and availability of peritoneal dialysis supplies, might have reduced the program's effectiveness.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Recruiting and enrolling participants

Northwell enrolled a total of 705 participants with Stage 4 or 5 CKD in Healthy Transitions, 141 percent of its original enrollment goal. Although about one-quarter of all participants disenrolled from the program when their disease progressed to end-stage renal disease (ESRD), most remained in the program until the end of the cooperative agreement. Disenrollment due to death, election of hospice care, and moving out of the geographic area was low.

Nurse care managers identified potential program participants from patients' charts, conferred with nephrologists about the patients' eligibility, and determined whether patients were appropriate for the program. As mentioned, the referral and enrollment of high-acuity CKD patients who were already on the brink of dialysis disrupted the planning and preventive aspects critical to the theory of action. This in turn would have limited the potential reductions in hospitalizations, ED use, and costs. For example, a kidney transplant is an expensive procedure requiring hospitalization and high short-term costs, but facilitating an earlier transplant, if successful, can prevent years of dialysis in the future.

Engaging participants

Engaging participants, particularly through home visits, was a critical component of the Healthy Transitions model. Nurse care managers believed that the home visit set the program apart from other disease management interventions. Nurses tried to follow up with participants by phone once a month. They called weekly for high-risk patients. The nurses scheduled annual follow-up home visits, unless they determined a need for another home visit sooner. Evidence from interviews, survey data, and awardee reports indicated that the awardee delivered the intervention services as intended. Patients' satisfaction levels routinely exceeded 95 percent on the awardee's semiannual patient surveys. Program leaders, nonclinician staff, and nephrologists interviewed during the study agreed that the program largely achieved its objectives of improving quality of care, reducing cost of care, and making CKD care more patient-centered. The awardee's internal reports showed an increase in the number of safe starts (that is, nonemergent initiation of dialysis) and a reduction in the use of central venous catheters, which in turn could reduce hospitalizations.

Characteristics of Medicare FFS participants

The awardee enrolled 705 participants, but only 203 beneficiaries (29 percent) of them were enrolled in Medicare FFS and met the claims-based eligibility criteria for the study from November 2014 through August 2017 (Appendix A, Table A.1). The study excluded 355 patients with missing enrollment data and 147 patients who were enrolled in Medicare Advantage, Medicaid, or private insurance, as well as beneficiaries who were not enrolled in Medicare for at least three months before and after enrollment.

The average age of this subset of participants was 75 (Table 3). Most were male (56 percent) and White (70 percent). Most participants were originally eligible for Medicare based on age (70 percent), whereas 27 percent qualified because of a disability. In addition, 20 percent were dually eligible for Medicare and Medicaid, slightly lower than the national average of 21 percent.[4] Overall, participants in the Healthy Transitions program were substantially less healthy and had a greater need for care than the general Medicare FFS population, as evidenced by the fact that the average hierarchical condition category risk score for participants was two-and-a-half times (2.5) the average score for Medicare FFS beneficiaries nationwide (1.0).

Consistent with their high needs, participants had high rates of Medicare expenditures and service use in the year before enrollment (Table 3). The total average per beneficiary per month Medicare payment during the baseline year was $2,226, which was substantially higher than the 2014 national average of $816. In addition, 34 percent of the participants had an ED visit, resulting in an annual rate of outpatient ED visits of 477 per 1,000 participants during the baseline year. Almost half (46 percent) of participants had a hospital admission in the baseline

[4] The national data here and in the next paragraph are from the Centers for Medicare & Medicaid Services. "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

year, leading to an annual rate of acute hospital admissions of 796 per 1,000 participants, compared with 282 stays per 1,000 beneficiaries nationally. Similarly, the 30-day unplanned readmission rate for participants (10 percent) was much higher than the national rate of 18 percent per discharge.

Table 3. Baseline characteristics of Medicare FFS participants

Characteristics	Medicare FFS participants (N = 203)
Demographics	
Age at enrollment, years	75
Younger than 65, %	13
65 to 74, %	32
75 to 84, %	37
85 and older, %	18
Male, %	56
White, %	70
Black, %	17
American Indian, Alaska Native, Asian/Pacific Island American, or other, %	9
Hispanic, %	2
Unknown, %	2
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	70
Disability insurance benefits	27
ESRD[a]	3
Medicare and Medicaid dual status, %	
Dual	20
HCC score[a]	
Mean	2.5
25th percentile	1.5
Median	2.3
75th percentile	3.3
Service use and expenditures in year before enrollment	
Any hospitalizations, %	46
Number of hospital admissions (per 1,000)	796
Any outpatient ED visits, %	34
Number of outpatient ED or OBS visits (per 1 ,000)	477
Proportion of beneficiaries with a readmission, %	10
Total Medicare expenditures ($ PBPM)	2,226
Acute inpatient expenditures ($ PBPM)	1,004

Sources: Mathematica's analysis of information from the awardee's program finder file and Medicare claims and enrollment data from November 2014 through August 31, 2018, as of December 2019.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

Table 3 (*continued*)

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; OBS = observation stay; PBPM = per beneficiary per month.

Challenges of estimating program impacts

A rigorous impact evaluation was not possible for the CKD program for two main reasons. First, there were too few Medicare FFS beneficiaries in the analysis sample to detect effects as large as 20 percent on outcomes such as hospitalization, ED use, or expenditures. Second, there were serious concerns about identifying a valid comparison group because claims data lacked information on two clinical eligibility criteria and program staff reported sociodemographic factors such as poor housing and cleanliness affected enrollment decisions (due to the program's home visit requirement). Selection bias was also a concern because most cases of CKD are not identified in claims until later stages of the disease. This implies that unobservable factors likely influence when patients engage in care for CKD, which would limit the comparability of the small treatment group and any comparison group constructed solely using claims. In addition, the evaluation could not examine quality measures (such as the SF-12 Physical Component Score, Mental Component Score, and KDQOL measures)—important intermediate outcomes that could be useful for identifying shorter-term effects.

CONCLUSIONS

Northwell was successful in implementing the Healthy Transitions care coordination and shared decision-making model for individuals with Stage 4 or 5 CKD living in eastern New York City and western and central parts of Long Island. Despite implementation challenges, the awardee exceeded its enrollment goals and was successful in enrolling and engaging providers, engaging patients, and delivering services. One notable concern, however, was that the program enrolled many people who were on the brink of dialysis. This timing disrupted the planning and preventive aspects critical to the theory of action and likely limited potential reductions in hospitalizations, ED use, and costs, at least during the short follow-up period when data for analysis would have been available. Several external factors could have affected program effectiveness as well, including patients' low health literacy and language barriers, the poor condition of patients' housing, and the lack of availability of peritoneal dialysis supplies. Due to the small sample size and an inability to replicate the eligibility criteria by using Medicare claims data , it was not possible to identify a comparison group that was similar to the intervention group at the time of enrollment into the program. Thus, it was not possible to conduct a rigorous impact evaluation of this program.

PROGRAM SUSTAINABILITY

After its award ended in February 2018, all but one of Northwell's sites continued the Healthy Transitions program. Internal funding from the Northwell Health System temporarily sustained the program at these sites. Northwell collaborated with the National Kidney Foundation to develop a payment model that combined a condition-specific, population-based payment with value-based incentives and penalties for nephrologists. Under this payment model, nephrologists would receive a monthly population-based payment for each patient who had pre-ESRD CKD. The payments would cover costs of care coordination and management, including patients' education and dietary assessments. The incentive payments would reward nephrologists for their quality performance related to AV fistula or graft placements, transplant listings, preemptive transplantation, initiation of home or outpatient hemodialysis or peritoneal dialysis, and conservative care management. Providers would be penalized for catheter placement, lack of hepatitis B and influenza vaccination, and anemia among patients who start hemodialysis. Although awardee leaders were optimistic that negotiations with a private company would generate enough funding to sustain the program for at least five years, they had not engaged any payers to fund the payment model by the end of the award.

Northwell's proposed payment model

Northwell collaborated with the National Kidney Foundation to develop a payment model that combined a condition-specific, population-based payment with value-based incentives and penalties for nephrologists. The payments would cover care coordination and care management costs. Providers would be rewarded or penalized for their performance on select quality measures.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants		**705**
Invalid or missing enrollment data, date of birth, or sex	355	350
Not enrolled in Medicare or could not be identified in Medicare enrollment files	90	260
Enrolled in Medicare Advantage, not enrolled in both Medicare Parts A and B, and insufficient FFS enrollment period at baseline	57	203
Final analytic sample		**203**

Sources: Mathematica's analysis of information from the awardee's program finder file and Medicare claims and enrollment data from November 2014 through August 31, 2018.

FFS = fee-for-service.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Regents of the University of California at San Diego

September 2020

James Reschovsky, Stephanie Judy, Theresa Feeley-Summerl, Bob Schmitz, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

REGENTS OF THE UNIVERSITY OF CALIFORNIA AT SAN DIEGO

The Regents of the University of California at San Diego (UCSD) used funds from Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the Heart Attack and Stroke Free-Zone (HSF-Z) program, an effort to improve care for patients at elevated risk for cardiovascular disease (CVD). The goals of HSF-Z were to reduce the incidence of heart attacks and strokes in San Diego County, thereby decreasing mortality rates and associated health care costs. UCSD led HSF-Z in partnership with 10 San Diego-area health systems and medical groups.[1] USCD received an HCIA R2 award in September 2014 and launched its program in January 2015.

Important issues for understanding the evaluation

- A rigorous impact analysis of this program was not feasible. The use of clinical information unavailable in claims data to determine eligibility and the latitude afforded implementing sites in designing their own recruitment strategies precluded constructing a comparison group.
- The evaluation of HSF-Z is descriptive; it uses self-reported data from the awardee's administrative files on 4,158 participants to assess adherence to prescribed drugs and changes in patients' risk factors between baseline readings and their most recent follow-up.
- One-third of HSF-Z participants did not engage in the program (that is, they did not meet at least three times with a health coach). Even among engaged participants, follow-up clinical values were often unavailable in administrative data.
- Because there is no comparison group, it is impossible to infer causality between the program and changes in patients' risk for CVD after enrollment.

HSF-Z intended to improve participants' health by raising their awareness of CVD; introducing evidence-based medications; and providing supportive, ongoing health coaching. Health coaches worked with physicians to ensure participants were put on appropriate, evidence-based medication bundles for hypertension, diabetes, and other conditions that can raise the risk of major cardiovascular events, such as strokes and heart attacks. They also encouraged adherence to the drug regimens and lifestyle changes among patients. In addition, HSF-Z conducted a small-scale pilot test of wireless blood pressure monitoring and provided community-wide efforts to educate people about the risks for CVD (the Be There program) and to educate physicians on best practices (the University of Best Practices). Table 1 provides a summary of the key HSF-Z characteristics.

[1] The 10 medical groups and health systems were Sharp Rees Stealy, the Scripps Foundation, the University of California at San Diego Family Medicine Group, Vista Community Clinic, Neighborhood Healthcare, Arch Health Partners, San Ysidro Health Center, the North Coast Family Medical Group, the University of California at San Diego Internal Medicine, and North County Health Services.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The Regents of the University of California at San Diego implemented HSF-Z to reduce the incidence of heart attacks and strokes in San Diego County, along with their associated health care costs and mortality rates.
Major innovation	The program provided supportive, ongoing health coaching to participants. Health coaches educated participants on CVD and its prevention, including evidence-based medication bundles and lifestyle changes. Community-wide education efforts also targeted the general public and providers.
Program components	• Care management through health coaches, including medication management • Engagement and support of participants and providers • Pilot program featuring wireless monitoring of blood pressure at home
Target population	The program sought to engage Medicaid, Medicare, and dually eligible beneficiaries who were at high risk for major adverse cardiac events—specifically a heart attack, stroke, or sudden death due to cardiovascular complications—and were either not on an evidence-based medication bundle or were on evidence-based medications but not achieving blood pressure control (suggesting the need for dosing review or medication adherence).
Participating providers	Ten medical groups or health systems operating in the San Diego area participated. Health systems implemented HSF-Z across multiple practice locations.
Total enrollment	4,158 participants enrolled (104 percent of the original program goal).
Level of engagement	About one-third of program participants were unengaged, measured as having fewer than three encounters with health coaches. Most of these unengaged participants withdrew or were removed from the program by implementing organizations (most often for not meeting with health coaches.
Theory of change or theory of action	Providing participants with a health coach and appropriate evidence-based medication will reduce the incidence of cardiovascular events, improve survival rates, and reduce overall health care costs.
Award amount	$5,820,416
Effective launch date	• Participating patients were first enrolled in January 2015. • The descriptive evaluation includes enrolled beneficiaries through May 2017.
Program setting	Provider-based (primary care physicians)
Market area	Urban, suburban
Market location	San Diego County
Target outcomes	• Decrease incidence of major adverse cardiac events • Lower participants' mortality rates • Enhance participants' experience with physicians and staff in physicians' offices • Decrease total Medicare expenditures • Decrease rate of ED visits • Increase percentage of participants adhering to medications
Evaluation design	The inability to identify a credible comparison group precluded a rigorous impact evaluation. Results are limited to a comparison of selected clinical outcomes among participants after versus before enrollment based on awardee data.
Payment model	The awardee proposed a per-beneficiary annual payment to pay for the HSF-Z program, with rates ranging from $620 for beneficiaries in Medicare Advantage plans to $182 for those in Medicaid managed care plans.

Table 1 (*continued*)

Program characteristics	Description
Sustainability plans	USCD did not continue support health coaching or case management at implementing sites after the end of the award, and sustainability plans varied across implementing sites.

CVD = cardiovascular disease; ED = emergency department.; HSF-Z = Heart Attack and Stroke Free-Zone.

It was not possible to conduct a rigorous impact evaluation of the HSF-Z program because the clinical values used to determine patients' eligibility (for example, blood pressure levels) were unavailable in claims, precluding construction of a comparison group and rigorous evaluation of this program. Instead, this evaluation uses HSF-Z's administrative data files provided by the awardee to describe participants and examine the changes in their cardiovascular risk indicators during the program. Table 2 summarizes the key features of the descriptive analysis.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	A rigorous impact evaluation of this program was not possible because the program eligibility criteria included information from clinical records that were not available in claims data. This included identifying beneficiaries who were at high risk for a major adverse cardiac event (such as heart attack, stroke, or sudden death due to cardiovascular complications) and were either not on an evidence-based medication bundle or were on evidence-based medications but not achieving blood pressure control. Instead, a descriptive analysis of self-reported program data was conducted to measures changes in indicators of cardiovascular risk among program participants and to identify the factors associated with those changes.
Intervention group for descriptive analysis	The descriptive analysis of demographic characteristics included all 4,158 participants. Assessments of changes in cardiovascular risk factors were restricted to engaged participants, defined as those with three or more visits with health coaches (2,825), who had both baseline and follow-up clinical values for the risk factor in HSF-Z records. The number with participants with clinical values in both the pre- and post-implementation periods varied by risk factor.
Limitations	Due to the problems noted above, the analysis cannot be used to make inferences about the impact of this program on Medicare costs or other program outcomes.

ECC = early childhood caries.

PROGRAM DESIGN AND ADAPTATION

UCSD originally partnered with eight San Diego area health systems and medical groups, with two additional organizations added in 2016. Within these organizations, individual practices implemented HSF-Z. The number of participating practices grew over time. The original target enrollment goal for all three years of the program was 4,000 participants.[2]

HSF-Z trained health coaches to use an Ask-Educate-Ask methodology that focused on asking about barriers to taking recommended medications or making lifestyle changes, educating patients about the benefits of medication adherence and blood pressure control, and asking

[2] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

patients about their next steps. Participating organizations designed their own approach to staffing health coach positions. The organizations used varying types of new or existing staff as health coaches, including registered nurses, certified health coaches, medical assistants, pharmacists, behavioral specialists, and care coordinators. The program intended weekly encounters during the month of enrollment and until the patient achieved medication adherence and blood pressure control. Afterward, encounters were to occur monthly with longer, more in-depth encounters to occur annually to gather information on clinical indicators, including laboratory values.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

HSF-Z experienced delays in its launch due to data use agreement issues and recruitment of patients was initially slower than expected. In the end, after extending the enrollment deadline from May 2016 to February 2017, HSF-Z recruited 4,158 participants, more than the goal of 4,000. Continued engagement of these participants proved more challenging, however. More than one-third of enrolled patients withdrew or were removed from the program by implementing organizations. Some withdrew for reasons unrelated to the program (for example, they moved to another city or changed medical groups). Others left the program because they met their health goals and did not see value in continued contact with health coaches. Implementing organizations removed participants primarily for not meeting regularly with their health coaches. In the face of declining patient interest in continued monthly contact with health coaches, particularly among those who met their health goals, program leaders relaxed requirements for monthly meetings in the third year of the cooperative agreement and several sites employed texting as a less burdensome means to maintain contact with patients.

Implications of program implementation for achieving program goals

- The initial slow recruitment of participants reduced the time available to observe program impacts.
- Participating medical groups had considerable latitude in how they recruited patients and implemented HSF-Z; information on what features or practices might have contributed to greater improvements in intermediate outcomes was unavailable.
- About half of enrolled patients (51 percent) exhibited both some engagement (they met with patient coaches three or more times) and provided at least some follow-up data for measuring changes in intermediate outcomes after enrollment.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Several program characteristics precluded identifying a comparison group and consequently rigorous evaluation. Although overall eligibility requirements remained unchanged, HSF-Z afforded implementing organizations considerable latitude in how they identified and prioritized eligible participants for recruitment, and allowed the organizations to modify their strategies throughout the implementation period. Moreover, patients' eligibility depended on past clinical history and clinical values (for example, blood pressure, cholesterol, and hemoglobin A1c levels) related to cardiovascular risk, information unavailable in claims data. Finally, it would take many years to fully observe and assess program impacts on the incidence of strokes and heart attacks and their attendant costs. For these reasons, the study was limited to a descriptive analysis of intermediate outcomes based on awardee-reported data under HSF-Z.

The evaluation used the program's administrative data populated by HSF-Z program staff at implementing organizations. These data include baseline information on participants obtained at enrollment as well as information on patients' clinical values and adherence with drug regimens obtained during subsequent encounters. Participants' self-reports and medical records provided this information, and the data could not be corroborated by independent sources. The analysis lacked information to assess changes in the wireless blood pressure monitoring pilot. Nor was it able to assess the community-wide education efforts, as these affected broader populations of patients and providers for which information was unavailable.

Enrolling and engaging participants

Health coaches frequently reported difficulties keeping participants engaged in the program. Since unengaged participants were less likely to have provided follow-up data than engaged participants, this analysis is restricted to patients who exhibited at least a minimal level of engagement in HSF-Z. This analysis defines engaged participants as those with three or more encounters with health coaches (either in-person or by telephone) during the award. HSF-Z program leaders used this same criterion in their evaluation of the program.[3] Among the 4,158 enrolled participants, 2,825 (68 percent) met this criterion. The number of participants with clinical values in both the pre- and post-implementation periods varied by risk factor.

[3] Be There San Diego. "Making San Diego a Heart Attack and Stroke Free Zone." Available at http://betheresandiego.org/storage/files/CMMI%20Report%20Final.pdf. UCSD also excluded those enrolled for fewer than 90 days, which excluded an additional 15 beneficiaries.

Characteristics of program participants

More than 70 percent of HSF-Z participants were age 65 or older and white. A similar proportion were enrolled in either Medicare Advantage or traditional Medicare. An additional 6 percent were dually eligible for Medicare and Medicaid. Overall, the characteristics of the 50 percent of program participants categorized as engaged (those with three or more health coach visits) and who provided both baseline and follow-up data on blood pressure was similar to that of all enrollees (Table 3).

Number of participants in study

Total participants	4,158
Eligible for evaluation (3+ encounters)	2,825
Have both baseline and follow-up data on:	
Blood pressure	2,070
Low-density lipoprotein cholesterol	816
Blood sugar control (HgA1c)	1,059
Body mass index	1,922
Adherence to prescribed statins	2,011

Table 3. Demographic characteristics of program participants and those who were engaged and had follow-up data for blood pressure

Demographic characteristics	All participants (N = 4,158)	Engaged participants with baseline and follow-up data[a] (N = 2,070)
Age group, %		
Younger than 50	2.0	1.2
50 to 64	26	23
65 to 74	50	52
75 or older	22	25
Female, %	54	56
Hispanic, %	30	29
Race, %		
White	73	75
Black or African American	6	6
Asian or Pacific islander	8	6
Other, unknown, or missing	14	13
Insurance coverage, %		
Medicare FFS	29	29
Medicare Advantage	41	45
Medicaid (MediCal)	24	20
Dual eligible beneficiaries	6	6
Missing	0.2	0.0

Source: Mathematica's analysis of awardee program data, September 2017.

[a] With three or more health coach visits and both baseline and follow-up blood pressure readings.

FFS = fee-for service.

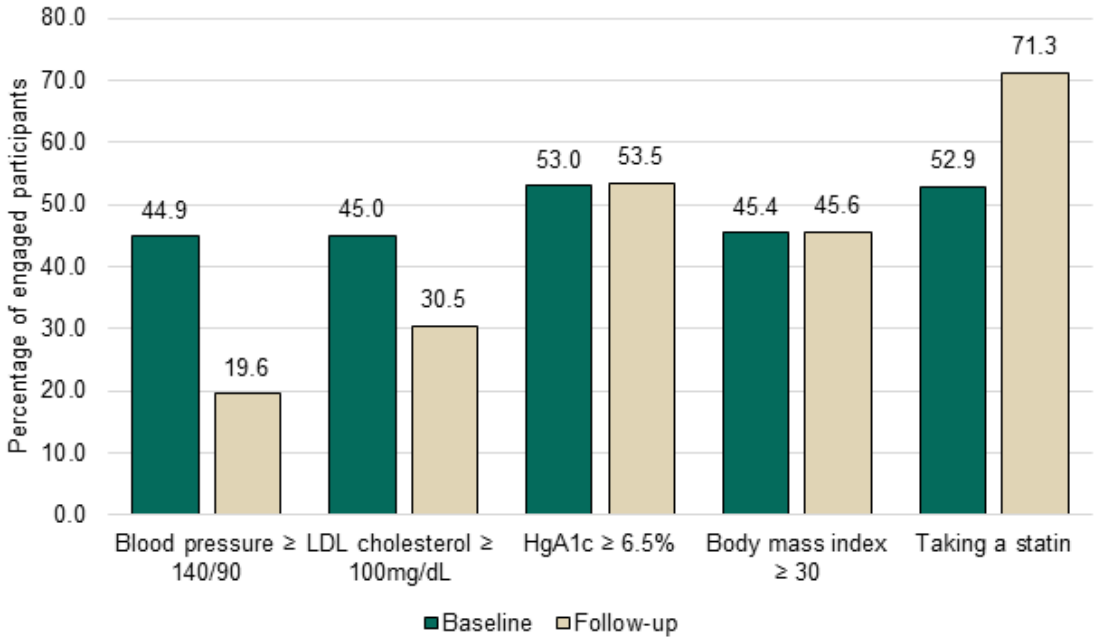
Analytic approach

This descriptive analysis assesses changes in intermediate risk factors for CVD among engaged participants after enrollment, as well as changes participants' adherence to recommended drug regimens. These assessments are limited to participants for whom participating organizations obtained both baseline and follow-up data on specific clinical values related to CVD risk and drug regimen compliance during annual visits.[4] Those who did not meet the definition of engagement, most of whom either withdrew from the program or who were removed by implementing organizations, were on average less likely to have achieved success in lowering CVD risk factors or following drug regimens. Therefore, these evaluation results are likely to show greater improvements in intermediate outcomes than if full data had been available on all engaged and unengaged HSF-Z participants.

DESCRIPTIVE RESULTS

Figure 1 shows HSF-Z participants' average change in four clinical scores indicating CVD risk: blood pressure, low-density lipoprotein (LDL) cholesterol, blood sugar control (as measured by hemoglobin A1c (HgA1c)), and body mass index (BMI), as well as increases in their use of statins, a class of drugs used to reduce cholesterol levels.

Figure 1. Changes in HSF-Z patients' cardiovascular risk factors and drug adherence among engaged participants with both baseline and follow-up data



Source: Mathematica analysis of awardee program data.
HSF-Z = Heart Attack and Stroke Free-Zone; LDL = low-density lipoprotein.

[4] The analyses excluded participants if their baseline values were collected more than six months before the date of enrollment and if follow-up data were collected fewer than three months after enrollment.

Among engaged participants for whom participating organizations obtained follow-up information, the greatest change was in reduced hypertension, defined as blood pressure in excess of 140/90, which fell to less than half the baseline level. The proportion of patients with elevated LDL cholesterol levels also fell substantially, by one-third. However, the proportion of HSF-Z's participants with elevated HgA1c levels among diabetics and the percentage of all participants who were obese remained virtually unchanged.

HSF-Z classified participants into three risk groups, depending on their age and history of high blood pressure, CVD, and diabetes. This analysis focuses on a single class of drugs, statins, which HSF-Z recommended for all three risk groups. The percentage of participants who took statins increased from 52.9 percent before enrollment to 71.3 percent at follow-up. However, at follow-up, only 59.3 percent of engaged participants reported taking statins on a daily basis, defined as six or seven times per week (data not shown).

CONCLUSION

The HSF-Z program aimed to use patient coaches to educate patients about CVD and engage them in their own self-care to increase the use of recommended prescription drugs and reduce risk factors associated with heart attacks and strokes. Among engaged patients, the two cardiovascular risk factors (blood pressure and cholesterol) directly associated with the drug bundles they promoted declined substantially, but two other indicators of CVD risk (hemoglobin A1c levels (among diabetics) and BMI) did not improve. Reductions in blood pressure and LDL levels were modest in size and observed only in subsets of participants. These results are for engaged patients, those likely to be most motivated to improve their health, even in the absence of the program. Because the evaluation could not determine whether these patients would have improved without the HSF-Z program, it is not possible to assess the extent to which the improvements observed are attributable to the HSF-Z program. However, the substantial increase in use of statins (as promoted by the program) among engaged participants suggests that the reductions in blood pressure and cholesterol observed were due at least in part to the program.

Limitations of evaluation

The analysis has several limitations. First, because the evaluation lacked a comparison group, no causal inferences can be drawn from the results. Second, the evaluation presents only pre-post comparisons of CVD risk indicators and medication compliance for program participants who met with health coaches at least three times and for whom there was follow-up data collected at an annual visit. As a result, the findings are prone to selection bias and do not represent what program effects would have been if evaluated for all patients, including eligible patients who refused the invitation to enroll in HSF-Z. The results are based on a subset of participants who are likely to be more motivated to improve their health than participants missing from the analysis. Third, HSF-Z patients' cardiovascular risk factors and drug adherence were measured by program staff and could have been subject to unconscious measurement bias. It was not possible to verify the accuracy of their results. Finally, changes in CVD risk factors varied by

implementing organizations. But because important information was unavailable on the operational protocols used by these organizations, no lessons are available from this study for other organizations that might want to implement a similar program to help patients reduce their risk for CVD.

PROGRAM SUSTAINABILITY

At the end of the award in August 2017, UCSD stopped operating its HSF-Z program. The awardee ended regular contact with its partner sites but learned that at least one site continued a modified version of the program. That site launched a diabetes prevention program that, like the original program, uses health coaches, but differs in its focus on patients with diabetes only and on education rather than medication adherence.

UCSD failed to implement its payment model with any payers by the end of its award period but continued to hold discussions with health plans after its award ended.

UCSD's proposed payment model

The awardee proposed a per-beneficiary annual payment to pay for the HSF-Z program. The payments would cover adults with hypertension, hyperlipidemia, and/or diabetes who were either newly diagnosed or had difficulty managing their condition(s). The covered services would include an assessment of social risk factors, connections to community resources, an assigned community health worker, and a health coach for individuals requiring more intensive support.

Based on potential savings from reduced cardiovascular events using the Archimedes modeling software, the awardee estimated annual payments per beneficiary as follows:

- $620 for Medicare Advantage plans
- $559 for commercial plans
- $182 for Medicaid plans

These amounts factor in a discount to recognize reduced potential savings from beneficiaries who enroll for just one year, but are not adjusted based on beneficiaries' characteristics and medical complexity, or for performance on quality or spending metrics.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Regents of the University of California at San Francisco

September 2020

Shivani Reddy, Stephanie Peterson, Shannon Heitkamp, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

REGENTS OF THE UNIVERSITY OF CALIFORNIA AT SAN FRANCISCO

The Regents of the University of California at San Francisco (UCSF), in partnership with the University of Nebraska Medical Center (Omaha), received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Dementia Care Ecosystem program. The goal of the program was to develop and test a new proactive model of dementia care that provided personalized support and education to local dementia patients and their caregivers. The target population consisted of Medicare fee-for-service (FFS) and Medicaid beneficiaries ages 45 and older with dementia and an identified primary family caregiver (dyads). The Dementia Care Ecosystem launched on March 31, 2015, and the intervention period funded by HCIA R2 ended on February 28, 2017. Table 1 summarizes the key characteristics of the program.

Important issues for understanding the evaluation

- The Dementia Care Ecosystem aimed to improve satisfaction with dementia care, reduce unnecessary emergency department and hospital use, and delay nursing home placement, leading to overall cost savings.
- This impact analysis is based on a randomized controlled trial in which participants were randomized 2:1 to the intervention and control groups.
- This impact analysis is based on 538 Medicare enrollees; of them, 358 were assigned to treatment and 180 to control status.
- Lower-than-expected enrollment, lower-need participants, and lack of engagement of primary care providers make it more difficult to detect program effects.

The awardee hypothesized that delivering telephone-based support and education to dementia patients and caregivers would reduce caregivers' burden, improve patients' and caregivers' satisfaction with dementia care, and enable caregivers to better support dementia patients in the community. Care team navigators (CTNs) served as the point of contact for dyads, providing resources related to disease and behavior management, caregiver support, and legal and financial planning. A multidisciplinary clinical team supervised the CTNs, provided guidance, and performed medication reviews. The goals of the program were to reduce the incidence of medical emergencies, prevent unnecessary emergency department (ED) and hospital use, and delay nursing home placement, leading to overall cost savings to the health care system and improved quality of life for patients and families.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	To develop, improve, and test a new model of dementia care that would address the unmet needs of patients and caregivers in the current FFS payment structure
Major innovation	Provide telephone-based supportive care and education for caregivers and beneficiaries, as well as medication consultation and support in planning future medical, financial, and legal decisions
Program components	• Patient navigation to provide support and education, link beneficiaries with resources, and triage clinical questions to the appropriate medical professionals • Patient and family education targeted to the patient's needs and stage of dementia; resources include legal, financial, and medical planning, as well as behavior management and safety planning • Care coordination with a multidisciplinary clinical team (nurse, pharmacist, and social worker) trained to supervise CTNs and intervene when specialized guidance for medical decision making is needed • Medication management using a computerized dashboard system, at enrollment and at the request of a clinician or CTN • Health information technology to support delivery of dementia education components, caregiver support resources, and medication management, in addition to CTN workflow management (for example, scheduling and data collection tools)
Target population	Medicare FFS or Medicaid beneficiaries ages 45 and older with a diagnosis of dementia and their caregivers, including underserved populations in California, Iowa, and Nebraska
Total enrollment	780 beneficiaries (37 percent of original enrollment goal of 2,100)
Level of engagement	Of the 553 beneficiaries randomized to the treatment group and included in our evaluation sample, 81 (29 percent) had an acuity score of high, indicating moderate to end-stage dementia requiring monthly contact with the CTN and routine consultation with the clinical team; the remaining treatment group beneficiaries received monthly or less frequent contact with the CTN and as-needed or limited consultation from the clinical team; the evaluation included all beneficiaries, regardless of level of engagement
Theory of change or theory of action	The awardee hypothesized that giving beneficiaries and caregivers personalized preventive care provided over the phone and supported by innovative technology should reduce the incidence of medical emergencies, prevent unnecessary ED and hospital use, and delay nursing home placement; these outcomes should result in overall cost savings to the health care system and improved quality of life for beneficiaries and families
Award amount	$9,990,848
Effective launch date	Program became operational on March 31, 2015
Program settings	CTNs primarily engaged beneficiaries and caregivers by telephone and occasionally in person
Market area	Urban, suburban, and rural (California, Iowa, and Nebraska); participants were recruited from areas near the awardee's institutions in San Francisco (California, n = 463, 59 percent) and Omaha (Nebraska and neighboring areas of Iowa, n = 317, 41 percent)
Target outcomes	• Improved caregiver perception of beneficiary's quality of life • Heightened caregiver satisfaction with services • Reduced caregiver burden • Reduced caregiver depression • Decreased ED visit rate and costs, hospitalization costs, ambulance use and costs, nursing facility costs, prescription drug costs, use of high-risk medications and other potentially inappropriate medications, and percentage of patients with an adverse drug event

Table 1 (*continued*)

Program characteristics	Description
Payment model	UCSF proposed two payment models: the first model was for non-risk-sharing organizations that featured FFS billing codes for chronic care management and advanced care planning; the second model was a value-based payment model that had not been developed at the end of the award
Sustainability plans	UCSF received a five-year R01 award from the National Institute on Aging to partially sustain the program at the awardee's site after the end of the cooperative agreement; the awardee reported plans to use the five-year award to collect longitudinal program data necessary to show the benefits of the program and engage payers; the University of Nebraska Medical Center had not secured funding to sustain the program by the end of the award, but was working with two potential partners to fund an expanded version of the program for patients with significant comorbidities

This impact analysis is limited to Medicare beneficiaries who participated in the randomized controlled trial (RCT) designed by the awardee, met evaluation claims and Medicare eligibility criteria, and to outcomes that could be measured with Medicare and survey data available for both treatment and control arms. All eligible beneficiaries were randomized to a treatment or control group at the time of enrollment. Enrollees randomized to the treatment group received services through the Dementia Care Ecosystem. Control group enrollees received usual care and completed surveys. Table 2 summarizes the key features of the evaluation.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a post-period cross-sectional model that compared the change in outcomes among treatment beneficiaries after enrollment relative to the change in outcomes over the same period among the randomized control group.
Treatment group for evaluation	The impact analysis relied on 538 Medicare FFS beneficiaries who were randomized to either the treatment or control group for the Dementia Care Ecosystem program from March 2015 to February 2017. The study sample excluded 112 randomized patients who either were not enrolled in Medicare or could not be identified in the Medicare enrollment database, and 122 beneficiaries who did not meet the claims-based eligibility criteria of the study, such as being enrolled in Medicare Parts A and B FFS with Medicare as primary payer. An additional 7 beneficiaries were excluded because they did not have 90 days of Medicare FFS eligibility during the baseline period and 1 was excluded because the beneficiary died within 30 days of randomization. Of the 538 beneficiaries, 358 were randomized to the treatment group.
Control group for evaluation	Of the 538 study beneficiaries, 180 were randomized to the control group.
Limitations	The small study sample makes it unlikely that statistically significant results would be obtained unless the true program effects were quite large.

PROGRAM DESIGN AND ADAPTATION

The Dementia Care Ecosystem had three key components: (1) patient navigation by the CTN; (2) care coordination, including medication management, by a multidisciplinary care team; and (3) health information technology (health IT).[1]

Patient navigation

The Dementia Care Ecosystem program provided care management and caregiver support using CTNs, who also linked participants with any needed community resources. The care management provided by the CTNs, who were unlicensed health providers, included telephonic support for dementia patients and their caregivers and triaging of more advanced needs to the Dementia Care Ecosystem clinical team. Caregivers also had access to educational resources for legal, financial, and medical planning, as well as dementia care delivery from the CTN.

Although the overall service delivery model was standardized in its core program elements—including the emphasis on the CTN being the point of contact to provide individualized caregiver support and education over the phone throughout the cooperative agreement period—the program team refined the model over time based on user inputs. In Year 3, this refinement involved developing the Dementia Care Ecosystem Lite program, which modified the frequency and intensity of CTN-delivered services for lower-need dyads. The Ecosystem Lite program was designed for patients and caregivers who requested fewer check-in calls, were not responsive to the CTNs' phone calls, or had low acuity scores. (For more information on acuity scores, see the section on participant enrollment, engagement, and activation.) Furthermore, the awardee modified the duration of services provided. Originally, the awardee had planned to include participants in the program indefinitely due to the natural progression of dementia as a disease. However, during the third program year, staff began to graduate program participants with low acuity scores, after they had at least one year of exposure to the intervention, due to constraints in available resources, including CTNs leaving the program to attend graduate school.

Care coordination and medication management

The multidisciplinary Dementia Care Ecosystem clinical team—consisting of a nurse, a pharmacist, and a social worker—intervened when participants needed specialized attention or CTNs needed guidance to make medical decisions. In addition, all participants received a medication review by the team pharmacist at enrollment. A clinician or CTN, or an automated alert from the computerized dashboard system used by program staff, could also trigger a medication review. This intervention component was relatively stable and was not modified over the course of the cooperative agreement. The program also intended to directly coordinate care with the primary care providers of participants to integrate care plans and medication changes into outpatient care, though these aspects faced difficulties as described later.

[1] The Third Annual Evaluation report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Health IT

Program staff used a clinical workflow management application (the dashboard) to support patient navigation, medication management, and other Dementia Care Ecosystem activities. The dashboard included scheduling and data collection tools. Staff developed and optimized the dashboard in Year 1 and, by Year 2, it was the standard way for the CTNs and other staff to document delivery of the core Dementia Care Ecosystem modules and coordinate care for each dyad among the care team.

The awardee did not implement two technology-driven components of the intervention: the caregiver portal and the functional monitoring module. The caregiver online portal aimed to provide asynchronous access to the CTN, a forum to engage with other caregivers, and a repository of educational materials. Competing demands on staff of both enrolling and providing the other program components limited the awardee's rollout of the portal, as well as lack of interest among most participants in using this technology. The functional monitoring module involved using smartphones and sensors in participants' homes to rapidly detect and respond to changes in functional status (that is, falls and unresponsiveness). The awardee rolled out the functional monitoring module only for a trial period in Nebraska. As with the caregiver module, the elderly participant population was not interested in engaging with the technology and CTNs were unable to accurately troubleshoot technological issues.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee delivered intervention services consistent with its planned approach of implementing an agile program, which was iterative and continually updated based on input from staff members, caregivers, and providers. Although the planned mode and quality of navigation care (that is, telephone support and education by CTNs) remained largely the same throughout the intervention, the frequency, intensity, and duration of the delivered services varied based on caregivers' and patients' needs and resource availability. The awardee's ability to recruit and hire qualified staff was for the most part successful, as was retaining most of them throughout the majority of the cooperative agreement. The main exception was loss of CTNs in Year 3 of the program. The awardee also provided effective, comprehensive training to staff in its service delivery protocol as well as other professional development and wellness (for example, stress reduction) trainings. After successfully recruiting and enrolling dyads in the program, the awardee engaged them in a timely manner, tailored services to the participants' needs by creating an acuity score to identify dyads who wanted less contact, and retained most of them for at least one year. Staff reported that caregivers really appreciated the relationship with the CTN and the access to the dementia specialists. In addition, telephone-based care enabled dyads that might normally be harder to reach in person to participate.

Throughout the cooperative agreement, the awardee struggled to recruit and enroll participants in the program. Initial recruitment strategies, such as relying on area agencies on aging to provide a large number of referrals, particularly from resource limited settings, produced only modest results. The awardee hired additional staff to bolster recruiting efforts. Strategies in Year 2 included attempting to recruit participants through a private senior care franchise and maximizing media coverage and outreach at local health and wellness events. In Year 3, the program relied heavily on two subspecialty referral sources internal to the awardee's health systems—a geriatric clinic at the University of Nebraska Medical Center and the UCSF Fresno Alzheimer and Memory Center—to meet modified enrollment targets. Most recruiting came from these internal referral sources, resulting in a disproportionately white, high-income, and English-speaking sample, rather than the low-income, ethnically diverse, underserved population that the awardee had intended to reach. As a result, participants had access to more services at baseline and less need for resources offered by the intervention, compared to the underserved population the awardee intended to recruit and believed would benefit more from the program.

Implications of program implementation for detecting impacts

- Final enrollment numbers were lower than originally expected, reducing the power to detect program effects.
- Dyads did not primarily come from low-income and underserved communities as anticipated. Enrolled dyads had fewer needs, leading to the creation of a Lite version of the intervention that might have made it harder to observe an intervention effect.
- Difficulty integrating the program with participants' primary care providers might have limited the impact of planned care coordination and medication reconciliation.

Furthermore, the Dementia Care Ecosystem was never fully integrated with primary care providers outside the networks of UCSF and the University of Nebraska Medical Center. Program staff found external providers difficult to engage due to their busy schedules and relative disinterest in the program. Multiple attempts by staff and program leadership to contact providers in many ways—such as by fax, email, and phone—were unsuccessful. As a result, CTNs worked with caregivers to empower them to engage with the dementia patient's provider directly.

ESTIMATING PROGRAM IMPACTS

Enrolling, engaging, and activating participants

Beneficiaries assigned to the treatment group received the intervention components described earlier, whereas control group beneficiaries received usual care and completed follow-up surveys. After dyads were successfully recruited and enrolled dyads in the program, the awardee engaged them in a timely manner, tailored services to the participants' needs, and retained most of them for at least one year. However, due to the higher socioeconomic status of recruited

participants, many dyads did not desire or require additional resources for dementia management. As a result, after the program started, the awardee created an acuity score for intervention group dyads and transitioned low-acuity dyads to Dementia Care Ecosystem Lite, a program that delivered less frequent and intense services. About 69 percent of intervention dyads had an acuity score of 2 or less, indicating monthly or less frequent contact by a CTN and intermittent or as-needed consultation by the clinical team. The impact analysis includes all treatment group members, including those with lower acuity scores, to ensure estimates are unbiased.

Study sample

In March 2015, the awardee began screening patients with dementia who were referred to the program and enrolling eligible beneficiaries. Final enrollment was substantially lower than initially planned because the initial recruitment strategies failed to generate a large number of enrollees, as described previously, and due to delays in staffing. All beneficiaries completed baseline questionnaires before assignment to a treatment or control group. The awardee then randomized 780 beneficiaries by dementia severity and by recruitment site in a 2:1 ratio to the treatment or control group. The evaluation sample was 538 Medicare FFS beneficiaries who were randomized to either the treatment or control group. The study sample excluded 112 randomized patients who either were not enrolled in Medicare or could not be identified in the Medicare enrollment database, and 122 beneficiaries who did not meet the claims-based eligibility criteria of the study, such as being enrolled in Medicare Parts A and B FFS with Medicare as primary payer. An additional 7 beneficiaries were excluded because they did not have 90 days of Medicare FFS eligibility during the baseline period and 1 was excluded because the beneficiary died within 30 days of randomization. The final evaluation sample consisted of 358 treatment group and 180 control group beneficiaries. (Appendix A, Table A.1 describes the identification of the analytic sample).

Characteristics of treatment and control group beneficiaries

A comparison of treatment and control group characteristics at baseline confirmed that the two groups are well balanced on most characteristics. (Table 3). The treatment group had 27 percent fewer ED visits than the control group in the year before enrollment, though this sizable difference was not statistically significant. Appendix B provides the full balance results measured during the 12 months before enrollment in the RCT.

Most of the sample was older than 75, and more than 86 percent of the sample was white. The average hierarchical condition category (HCC) risk score of participants was 1.3 for the treatment group and 1.4 for the control group, indicating that their expected Medicare expenditures were at least 30 percent higher than the general Medicare FFS population. Staff used the Quick Dementia Rating System (QDRS) to assess dementia severity at baseline, with average QDRS scores for the sample indicating mild to moderate dementia for both groups.

The rate of acute care hospitalizations was similar for the two groups during the baseline year. The rate of ED visits for the treatment group differed substantially from that of the control group, though the difference was not statistically significant. The rate of 590 per 1,000 participants per year for the treatment group was 27 percent lower than the rate of 749 per 1,000 for the control group. This disparity is likely attributed to the fact that the control group contains only 180 sample members and a higher proportion of control beneficiaries in the top 5 percent of the distribution have higher numbers of ED visits compared with treatment beneficiaries. Average spending per beneficiary per month (PBPM) among the treatment group was $1,000 compared to $1,067 for the control group. The regression model included these variables control variables to adjust for these baseline differences and to improve the precision of the estimates.

Table 3. Baseline characteristics of treatment and control group beneficiaries

Measure	Treatment (N = 358)	Control (N = 180)
Demographics		
Age at enrollment (continuous)	79	78
Age group, %		
Younger than 65 years	2	4
65 to 74 years	30	34
75 to 84 years	42	34
85 years and older	26	27
Female, %	55	54
Race/ethnicity, %		
White	87	86
Black	3	4
Hispanic	3	3
Other race/ethnicity	5	6
Health status		
HCC score (average)[a]	1.3	1.4
Dementia severity[b]	12	13
Service use and expenditures in the year before enrollment		
Hospital stays, per 1,000 beneficiaries	283	290
ED/observation visits, per 1,000 beneficiaries	590	749
Primary care visits in ambulatory settings, per 1,000 beneficiaries	5,142	5,504
Specialist visits in ambulatory settings, per 1,000 beneficiaries	8,195	7,973
Total Medicare expenditures ($ PBPM)	$1,000	$1,067

Source: Mathematica's analysis of information from awardee's randomization file and Medicare claims and enrollment data from March 31, 2014, to February 28, 2017.

Notes: The baseline period covers the 12-month period prior to the randomization date.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

None of the differences between treatment and control groups in any of the baseline characteristics was statistically different from zero at the 0.10 level, 2-tailed test.

Full balance results are presented in Appendix B.

Table 3 (*continued*)

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

[b] Dementia severity was assessed using the Quick Dementia Rating System, a 10-item questionnaire with scores ranging from 0 to 30. Higher QDRS scores correspond to increasing cognitive impairment: 0–1 (normal), 2–5 (mild cognitive impairment), 6–12 (mild dementia), 13–20 (moderate dementia), 21–31 (severe dementia). Source: Galvin, J.E. "The Quick Dementia Rating System (QDRS): A Rapid Dementia Staging Tool." Alzheimer's Dementia (Amst), vol. 1, no. 2, 2015, pp. 249–259. doi:10.1016/j.dadm.2015.03.003).

ED = emergency department; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact estimates were obtained from a regression of key outcomes on enrollees' characteristics and baseline expenditures and service use and cover the first 12-month period after enrollment, which occurred from March 31, 2015, to February 28, 2017. The regression should control for the sizeable difference between the treatment and control groups on ED visits in the year before enrollment, as well as the smaller differences on other covariates. The core study outcomes were obtained from Medicare claims. In addition to claims-based outcomes, the analysis included a dementia-specific quality of life outcome collected in surveys administered to both treatment and control groups.[2] Appendix A describes in detail the statistical models used to estimate the effects of the program. Appendix C contains a detailed description of the statistical model used to estimate the effects of the program. The program impacts for UCSF were also estimated using a Bayesian approach, and presented in Appendix D.

IMPACT RESULTS

While the program showed promising results – with the beneficiaries in the intervention group having somewhat (about 10 percent) fewer hospitalizations and ED visits, and slightly (5 percent) lower expenditures, than those in the control group, these results were not statistically significant and may not be due to the program (Table 4). The small study sample makes it unlikely that statistically significant results would be obtained unless the true (unobserved) program effects were quite large. There were no meaningful differences in estimated rates of specialist care visits nor in the dementia-specific quality-of-life measure. The impact results were not sensitive to the length of the enrollee's baseline period nor to the trimming of outliers.

[2] The awardee collected additional caregiver measures, such as burden and depression. However, more than 25 percent of one-year follow-up surveys were missing, precluding analysis.

Table 4. Estimated impact of the Dementia Care Ecosystem on selected outcomes

	Full RCT group (N = 538) Estimated impact over a 12-month follow-up period
Total Medicare expenditures ($ PBPM)	
Impact	-$81
Percentage impact	-5%
p-value	0.65
Hospital stays, per 1,000 beneficiaries	
Impact	-46
Percentage impact	-10%
p-value	0.52
ED or observation visits, per 1,000 beneficiaries	
Impact	-63
Percentage impact	-9%
p-value	0.53
Primary care visits in ambulatory settings, per 1,000 beneficiaries	
Impact	-151
Percentage impact	-3%
p-value	0.69
Specialist visits in ambulatory settings, per 1,000 beneficiaries	
Impact	115
Percentage impact	2%
p-value	0.88
Dementia-specific quality of life[a]	
Impact	0.1
Percentage impact	< 1%
p-value	0.8

Source: Mathematica's analysis of information from awardee's randomization file and Medicare claims and enrollment data from March 31, 2015, to February 28, 2017.

Notes: Impact estimates are based on the regression-adjusted difference between the randomized treatment and control group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post-period minus the impact estimate). Appendix C presents full impact estimates. Appendix D shows the results from the Bayesian analysis.

[a] The awardee administered the Quality of Life–Alzheimer's Dementia measure, a 13-item measure that rates each item on a 4-point scale (1 indicates poor, 4 indicates excellent) and total score ranges from 13 to 52. An impact estimate of 0.1 is not clinically significant.

*Significantly different from zero at the .10 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month.

CONCLUSION

It is unclear whether these favorable estimates are due to true effects of the program, or to chance. The estimates are imprecise, due to the small sample sizes. Even if the program's (unknown) true effect was large (say, a 20 percent reduction in hospitalizations), it is highly unlikely that the estimate in a sample of this size would be statistically significant. Large effects are unlikely due to the relatively low-acuity patient population and inability to integrate the intervention with primary care providers.

Main findings from impact evaluation

- Patients enrolled in the Dementia Care Ecosystem program had somewhat fewer hospitalizations and ED visits and lower expenditures than the comparison group. However, the statistical evidence that this lower resource use was attributable to the program was weak, due to the small sample size.
- The program's inability to recruit a high-needs population of dementia patients and caregivers, and low engagement of primary care physicians, hindered its capacity to improve outcomes.

The awardee intended to enroll a low-income, ethnically diverse population who would have high unmet medical and social needs and benefit more from the Dementia Care Ecosystem. In practice, however, the final recruited sample had fewer needs than expected and requested less intervention contact. About 69 percent of intervention dyads had only monthly or less frequent contact by a CTN and intermittent or as-needed consultation by the clinical team, making it less likely for the program to have an effect. Finally, the program was never fully integrated with primary care providers. Outpatient primary care providers might have implemented the Dementia Care Ecosystem medication reviews and care plans designed to reduce the incidence of medical emergencies and prevent unnecessary ED and hospital use only sporadically into chronic care management.

Limitations of evaluation

As noted earlier, the small sample size greatly limited the ability to detect program effects, especially given that the lower-than-planned acuity of the patients served likely attenuated true effects. Also, the delayed recruitment of dyads and graduation of participants after 12 months of follow-up reduced the number of beneficiaries with longer-term exposure to the program. The lack of participants with longer-term exposure to the intervention makes it difficult to assess the sustainability of program effects, and to detect effects on outcomes on ED and hospitalization rates and associated expenditures. Attempting to study participants with more than 12 months of follow-up would have resulted in far fewer beneficiaries and even less power to detect effects. In addition, according to the intervention's theory of action, dementia support was primarily directed to the participants' caregivers. There was insufficient follow-up data on caregivers' burden, self-efficacy, and depression, measures that could have elucidated impacts on these intermediate outcomes. Such data would help identify promising features of the intervention.

PROGRAM SUSTAINABILITY

After its award ended in February 2018, UCSF reported using a five-year award from the National Institute on Aging to continue offering program services to some dyads previously enrolled in the program, while graduating others. The dyads that remained in the program included those who reported high caregiver burden scores at baseline. The awardee had planned to enroll a minimum of 21 Latino dyads in 2018 using funding from the five-year award.

UCSF's proposed payment model

UCSF originally proposed two separate payment models. The first was for non-risk sharing organizations like theirs, which featured FFS billing codes for the chronic care management and advanced care planning codes (specifically, G0505 code for care planning and dementia). The second would be a value-based payment model that was not yet developed.

UCSF remained hopeful that participating sites would be able to continue the program after the NIH award ends, based on progress with its FFS payment approach. The awardee reported helping two new clinics set up new billing mechanisms to use the FFS payment approach, which will generate funds to help the clinics sustain the program. In addition, the awardee hoped that participating sites and payers might be inclined to continue and fund the program beyond the NIH award, based on the awardee's early internal findings that suggested the program reduced cost and use.

At the time of this report, the awardee had not progressed with developing its proposed value-based payment model. It had not updated the payment model design, analyzed additional data beyond year one to determine specific costs, or reached agreements with payers, including the Medicare Next Generation Accountable Care Organization (ACO) Allina Health with which the awardee had originally partnered to implement the program. The awardee may continue developing its value-based payment model after claims data become available to it, which it plans to analyze in addition to program survey data.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

In the randomized controlled trial (RCT) approach, the awardee randomly assigned eligible beneficiaries to either a treatment or control group. Program effects were estimated using a regression model of the following form:

(1) $$Y_{it} = \alpha + \theta_t Treatment_i + \beta' X_i + \varepsilon_{it}$$

where Y_{it} is the outcome of individual *i* in period *t* (for example, total monthly Medicare expenditures during the *t*-th time period since he or she enrolled); α is a constant term; $Treatment_i$ is an indicator for whether the individual is assigned to the group that received program services; X_i are beneficiary characteristics including age, gender, race, original reason for Medicare eligibility, Medicare and Medicaid dually eligible status, hierarchical condition category (HCC) score, recruitment site indicator (California or Nebraska), dementia severity, and other pre-enrollment characteristics, including baseline values of outcome measures, total expenditures, and rates of hospitalization and outpatient emergency department (ED) visits including observational stays. ε_{it} is a random disturbance term.

Equation (1) was estimated with an indicator for each period used to obtain an estimate of all parameters for each six-month period $T = \{1, 2, \cdots P\}$ following the date of enrollment. The key parameter of interest is θ_t, which measures the impact of the program in participants' *t*-th period after enrolling. Thus, the model can be used to assess how program impacts vary with enrollees' length of exposure to the program. There was no significant difference in results by period so this report presents results for the total follow-up period of 12 months.

Appendix A of Volume I of this report provides details on the standard set of outcomes. In addition to claims-based outcomes, the model also estimated over a dementia-specific quality-of-life outcome. The awardee administered a quality-of-life survey, the Quality of Life–Alzheimer's Dementia measure (QOL-AD), a 13-item measure that focuses on domains thought to be important in cognitively impaired older adults. Each item is rated on a 4-point scale, where 1 indicates poor, 4 indicates excellent, and total score ranges from 13 to 52. Survey measures completed by caregivers rating participants' quality of life were also used because more than half of the participants were unable to complete the survey themselves due to cognitive impairment. The response rate for the baseline survey was 100 percent. The follow-up survey was completed by 389 caregivers in the analytic sample (response rate of 72 percent). For beneficiaries missing follow-up surveys because they had died before 12 months of follow-up (n = 36), the lowest possible QOL-AD score (13) to include these beneficiaries in the analytic sample was imputed.[3]

[3] Jönsson, Linus, Neils Andreasen, Lena Kilander, Hikka Soininen, Gunhild Waldemar, Harald Nygaard, Bengt Winblad, Maria Eriksdotter Jönhagen, Merja Hallikainen, and Anders Wimo. "Patient- and Proxy-Reported Utility in Alzheimer Disease Using the EuroQoL." *Alzheimer Disease and Associated Disorders,* vol. 20, 2006, pp. 49-55.

Table A.1 shows the how the analytic sample for this study was defined. It lists the reasons why participants were excluded and the number of participants excluded for each reason.

Table A.1. Identification of final sample for impact analysis for UCSF

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through February 2017		**780**
Enrollees not linked to the Medicare enrollment database	112	668
Enrollees dropped because not enrolled in Parts A and B	21	647
Enrollees dropped because enrolled in Medicare Advantage	91	556
Enrollees dropped because Medicare is not primary payer	10	546
Enrollees dropped because lacks 90 days of FFS enrollment during baseline	7	539
Enrollees dropped because died within 30 days of enrollment	1	538
Final analytic sample		**538**
Randomized to intervention group		358
Randomized to control group		180

Source: Mathematica's analysis of information from Medicare enrollment data from March 2014 through March 2017.

Appendix B

Results from balance assessment of treatment and control groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used to assess balance. The table displays the weighted means of baseline characteristics for the 358 treatment beneficiaries and the 180 control beneficiaries used in the impact analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable, which was calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than .10 are generally considered evidence of good balance on assessed variables. The variables include demographic characteristics (age, gender, and race); state of intervention site; Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category [HCC] score, dementia severity, and dementia-specific quality of life; Medicare expenditures in total and by type of service; and service use. The variables are measured over the 12 months before enrollment in the intervention.

The table also shows the results of the equivalency-of-means tests. The equivalence test p-values are the greater of two one-sided weighted t-test p-values, which assesses whether the control group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Overall, the matching achieves good balance on baseline sociodemographic and health status characteristics and prior utilization and Medicare expenditures, but there are a few exceptions. The vast majority of baseline characteristics are less than the .10 threshold. Standardized differences for four variables fall between 0.10 and 0.25 , with the means for each being lower for the treatment group than for the control group (acute care expenditures, skilled nursing facility expenditures, emergency department [ED] and observation visits per 1,000 beneficiaries, and percentage with any outpatient ED visit). The differences between treatment and control beneficiaries are not statistically significant, but the standardized differences are somewhat higher than the general goal that was set of having standardized differences within a .10 boundary. Other differences in means fall within the .10 threshold, but are large enough to raise potential concerns: treatment group members had 20 percent more hospitalizations than control group members in the three months before enrollment, and were more likely to be age 75 or older (68 versus 61 percent). Nonetheless, the two groups were much closer on number of hospitalizations for the year before enrollment and average age. Thus, this randomized group of intervention and treatment beneficiaries yields credible estimates of program effects, but the regression model also controls for the differences between the two groups on baseline characteristics. Finally, an omnibus test was performed in which the null hypothesis is that the treatment and control groups are balanced across all the covariates. The results are used to assess the closeness of fit between the treatment and control groups on key characteristics likely to be associated with study outcomes. The result of the omnibus test is not significant, meaning that joint hypothesis that treatment–control differences in means for all variables are zero cannot be rejected.

Table B.1. Baseline characteristics of treatment and control groups for UCSF

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	79 (0.42)	78 (0.70)	0.53 (0.80)	< +/-1	0.07	0.50	0.03
Age: younger than 65, %	2.2 (0.78)	4.4 (1.5)	-2.2 (1.7)	-99	-0.15	0.15	0.19
Age: 65 to 74, %	30 (2.4)	34 (3.5)	-4.3 (4.3)	-14	-0.09	0.31	0.05
Age: 75 to 84, %	42 (2.6)	34 (3.6)	7.5 (4.3)	18	0.15	0.10	0.13
Age: 85 and older, %	26 (2.3)	27 (3.3)	-0.97 (4.0)	-3.7	-0.02	0.81	< 0.01
Female, %	55 (2.6)	54 (3.7)	1.1 (4.5)	2.1	0.02	0.80	< 0.01
White, %	87 (1.8)	86 (2.6)	1.0 (3.1)	1.2	0.03	0.74	0.01
Black, %	3.4 (0.95)	3.9 (1.4)	-0.54 (1.7)	-16	-0.03	0.75	0.01
Hispanic, %	2.8 (0.87)	3.3 (1.3)	-0.54 (1.6)	-19	-0.03	0.73	0.01
Other, %	5.0 (1.2)	6.1 (1.8)	-1.1 (2.1)	-22	-0.05	0.60	0.02
Unknown, %	1.7 (0.68)	0.56 (0.56)	1.1 (0.90)	67	0.09	0.28	0.01
Dual eligibility status, %							
Dually eligible for Medicare and Medicaid	15 (1.9)	12 (2.4)	3.1 (3.1)	20	0.09	0.33	0.03
Health status							
HCC score[a]	1.31 (0.05)	1.37 (0.08)	-0.06 (0.10)	-4.9	-0.07	0.48	0.04
Caregiver on patient's quality of life, Alzheimer dementia score[b]	33 (0.32)	33 (0.44)	-0.35 (0.55)	-1.1	-0.06	0.52	0.02
Quick Dementia Rating System score, baseline[c]	12 (0.33)	13 (0.51)	-0.97 (0.63)	-8.2	-0.16	0.10	0.18

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Control mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Medicare expenditures ($ PBPM)							
Total expenditures, 12 months before enrollment	1,000 (78)	1,067 (152)	-68 (178)	-6.8	-0.05	0.66	0.05
Total expenditures, 3 months before enrollment	983 (125)	928 (150)	55 (200)	5.6	0.02	0.79	< 0.01
Acute inpatient expenditures, 12 months before enrollment	226 (30)	301 (83)	-75 (91)	-33	-0.13	0.30	0.23
Skilled nursing facility expenditures, 12 months before enrollment	80 (20)	159 (47)	-79 (52)	-98	-0.21	0.07	0.38
Service use (per 1,000 beneficiaries)							
Total hospitalizations	283 (31)	290 (52)	-7.1 (62)	-2.5	-0.01	0.90	0.01
Total hospitalizations, 3 months before enrollment	279 (63)	222 (75)	57 (103)	20	0.05	0.58	0.01
Total ED or observation visits	590 (52)	749 (85)	-158 (100)	-27	-0.16	0.10	0.19
Total ED or observation visits, 3 months before enrollment	704 (104)	689 (121)	15 (157)	2.1	0.01	0.93	< 0.01
Primary care visits, ambulatory setting	5,142 (222)	5,504 (314)	-362 (379)	-7.0	-0.09	0.35	0.04
Primary care visits, ambulatory setting, 3 months before enrollment	5,307 (286)	5,378 (410)	-71 (490)	-1.3	-0.01	0.89	< 0.01
Specialty care visits, ambulatory setting	8,195 (373)	7,973 (558)	223 (685)	2.7	0.03	0.73	0.01
Number of beneficiaries	358	180					

Source: Mathematica's analysis of information from awardee's randomization file and Medicare claims and enrollment data from March 31, 2014, to February 28, 2017.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the adjusted difference and the treatment group standard deviation. p-values come from a weighted two-sample t-test; equivalence test p-values are the greater of the two one-sided weighted t-tests of whether the true treatment-comparison difference exceeded 0.25 standard deviations of the variable.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

[b] The awardee administered the Quality of Life–Alzheimer's Dementia measure (QOL-AD), a 13-item measure where each item is rated on a 4-point scale (1 indicates poor, 4 indicates excellent) and total score ranges from 13 to 52. An impact estimate of 0.1 is not clinically significant.

Table B.1 (*continued*)

[c] Dementia severity was assessed using the Quick Dementia Rating System, a 10-item questionnaire with scores ranging from 0-30. Higher QDRS scores correspond to increasing cognitive impairment: 0-1 (normal), 2-5 (mild cognitive impairment), 6-12 (mild dementia), 13-20 (moderate dementia), 21-31 (severe dementia). Source: Galvin, J.E. "The Quick Dementia Rating System (QDRS): A Rapid Dementia Staging Tool." Alzheimer's Dementia (Amst). vol. 1, no. 2, 2015, pp. 249–259. doi:10.1016/j.dadm.2015.03.003).

ED = emergency department; HCC = hierarchical condition category; SE = standard error; PBPM = per beneficiary per month; UCSF = University of California San Francisco.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Table C.1 shows the impact estimates over one year of intervention follow-up. The models were estimated over Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percent impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. Impact estimates that are statistically different from zero at the .01, .05, and .10 levels, using a two-tailed test, are indicated with one, two, or three asterisks, respectively. Table C.2. displays sensitivity analyses using two years of baseline data and topcoding at the 98th percentile. No significant differences were observed from the main impact results. There were also no significant differences between the semiannual and 12-month follow-up models.

Table C.1. Estimated impact of the UCSF intervention on select Medicare FFS expenditures (dollars PBPM) and use measures during a 12-month follow-up period

	All beneficiaries				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value
Total expenditures ($ PBPM)					
Baseline year	1,008	1,074			
Months 1–6	1,361	1,405	-44 (223)	-3.1%	0.84
Months 7–12	1,462	1,584	-122 (233)	-7.7%	0.60
Months 1–12	1,410	1,491	-81 (178)	-5.4%	0.65
Acute inpatient expenditures ($ PBPM)					
Baseline year	227	304			
Months 1–6	366	416	-50 (121)	-12%	0.68
Months 7–12	473	488	-15 (119)	-3.2%	0.90
Months 1–12	418	451	-33 (91)	-7.3%	0.72
Outpatient expenditures ($ PBPM)					
Baseline year	199	158			
Months 1–6	218	220	-1.4 (62)	< 1%	0.98
Months 7–12	164	174	-10 (45)	-5.9%	0.82
Months 1–12	192	198	-5.7 (44)	-2.9%	0.90
Professional services ($ PBPM)					
Baseline year	255	256			
Months 1–6	248	266	-18 (36)	-6.7%	0.63
Months 7–12	252	273	-21 (43)	-7.7%	0.62
Months 1–12	250	269	-19 (37)	-7.2%	0.61
SNF expenditures ($ PBPM)					
Baseline year	81	159			
Months 1–6	176	160	16 (69)	10%	0.81
Months 7–12	240	208	32 (83)	15%	0.70
Months 1 - 12	207	183	24 (57)	13%	0.67

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value
Hospital stays, per 1,000 beneficiaries					
Baseline year	286	294			
Months 1–6	347	399	-52 (98)	-13%	0.59
Months 7–12	456	495	-39 (98)	-7.9%	0.69
Months 1–12	399	445	-46 (72)	-10%	0.52
ED or observation visits, per 1,000 beneficiaries					
Baseline year	518	672			
Months 1–6	619	769	-150 (129)	-19%	0.24
Months 7–12	675	644	31 (130)	4.8%	0.81
Months 1–12	646	709	-63 (100)	-8.9%	0.53
Primary care visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	5,182	5,462			
Months 1–6	5,170	5,323	-153 (435)	-2.9%	0.72
Months 7–12	5,324	5,474	-150 (449)	-2.7%	0.74
Months 1–12	5,244	5,396	-151 (377)	-2.8%	0.69
Specialist visits in ambulatory settings, per 1,000 beneficiaries					
Baseline year	8,214	7,998			
Months 1–6	7,562	7,532	29 (759)	< 1%	0.97
Months 7–12	7,330	7,121	209 (841)	2.9%	0.80
Months 1–12	7,450	7,334	115 (744)	1.6%	0.88
Percentage of beneficiaries with any hospitalization stay in time period					
Baseline year	22	20			
Months 1–6	14	15	-0.94 (3.3)	-6.5%	0.77
Months 7–12	18	20	-2.1 (3.6)	-11%	0.55
Months 1–12	27	30	-2.8 (4.2)	-9.3%	0.51
Percentage of beneficiaries with any ED or observation visits in time period					
Baseline year	36	42			
Months 1–6	27	29	-2.8 (4.1)	-9.5%	0.49
Months 7–12	26	27	-1.1 (4.2)	-4.2%	0.79
Months 1–12	41	45	-3.6 (4.6)	-8.1%	0.43
Sample sizes					
Number of beneficiaries					
Baseline year	358	180			
Months 1–6	358	180			
Months 7–12	340	171			
Months 1–12	358	180			

Source: Mathematica's analysis of information from awardee's randomization file and Medicare claims and enrollment data from March 31, 2014, to February 28, 2017.

Table C.1 (*continued*)

Note: Impact estimates are based on the regression-adjusted difference between the randomized treatment and control group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post-period minus the impact estimate).

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SE = standard error; SNF = skilled nursing facility, UCSF = University of California San Francisco.

Table C.2. Estimated impact of the UCSF intervention on Medicare FFS expenditures (dollars PBPM) and use measures during a 12-month follow-up period, based on 2-year baseline and topcoding at 98th percentile

	Results from using 2-year baseline period					Results from topcoding data at 98th percentile[a]				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value
Total payments ($ PBPM)[b]										
Baseline year	908	900				1,008	1,030			
Months 1–6	1,361	1,421	-60 (221)	-4.2%	0.79	1,254	1,295	-41 (186)	-3.2%	0.82
Months 7–12	1,462	1,595	-133 (233)	-8.3%	0.57	1,321	1,527	-206 (209)	-13%	0.32
Months 1–12	1,410	1,504	-94 (177)	-6.3%	0.59	1,352	1,462	-109 (169)	-7.5%	0.52
Acute inpatient payments ($ PBPM)										
Baseline	197	258				225	252			
Months 1–6	366	410	-44 (121)	-11%	0.72	267	286	-20 (75)	-6.8%	0.79
Months 7–12	473	484	-11 (120)	-2.3%	0.93	330	397	-66 (82)	-17%	0.42
Months 1–12	418	446	-28 (91)	-6.3%	0.76	350	387	-37 (70)	-9.5%	0.60
Outpatient payments ($ PBPM)										
Baseline	196	150				170	146			
Months 1–6	218	230	-12 (62)	-5.1%	0.85	170	167	2.9 (29)	1.7%	0.92
Months 7–12	164	182	-18 (43)	-10%	0.67	141	168	-27 (30)	-16%	0.37
Months 1–12	192	207	-15 (42)	-7.3%	0.72	163	184	-21 (29)	-11%	0.46
Professional services ($ PBPM)										
Baseline	249	235				242	233			
Months 1–6	248	270	-22 (36)	-8.1%	0.55	241	238	2.3 (25)	< 1%	0.93
Months 7–12	252	275	-24 (42)	-8.7%	0.57	241	238	2.5 (25)	1.0%	0.92
Months 1–12	250	273	-23 (37)	-8.3%	0.54	245	240	5.1 (23)	2.1%	0.82
SNF payments ($ PBPM)										
Baseline	60	109				78	142			
Months 1–6	176	161	15 (69)	9.6%	0.82	135	123	12 (48)	9.4%	0.81
Months 7–12	240	210	31 (83)	15%	0.71	173	147	25 (56)	17%	0.65
Months 1–12	207	184	23 (56)	12%	0.68	187	174	13 (51)	7.7%	0.79

Table C.2 (*continued*)

	Results from using 2-year baseline period					Results from topcoding data at 98th percentile[a]				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value
Hospital stays, per 1,000 beneficiaries										
Baseline	252	258				286	288			
Months 1–6	347	396	-49 (97)	-12%	0.61	319	339	-20 (82)	-5.8%	0.81
Months 7–12	456	493	-37 (97)	-7.5%	0.71	415	459	-45 (87)	-9.7%	0.61
Months 1–12	399	442	-43 (72)	-9.7%	0.55	376	412	-36 (65)	-8.8%	0.58
ED or observation visits, per 1,000 beneficiaries										
Baseline	531	666				504	661			
Months 1–6	619	763	-143 (127)	-19%	0.26	572	732	-160 (106)	-22%	0.13
Months 7–12	675	634	41 (129)	6.5%	0.75	624	587	37 (112)	6.3%	0.74
Months 1–12	646	701	-55 (99)	-7.8%	0.58	613	692	-79 (91)	-11%	0.39
Primary care visits in ambulatory settings, per 1,000 beneficiaries										
Baseline	5,132	5,041				5,154	5,413			
Months 1–6	5,170	5,314	-144 (433)	-2.7%	0.74	5,103	5,159	-55 (397)	-1.1%	0.89
Months 7–12	5,324	5,465	-141 (450)	-2.6%	0.75	5,148	5,287	-138 (405)	-2.6%	0.73
Months 1–12	5,244	5,387	-143 (377)	-2.6%	0.70	5,204	5,345	-141 (362)	-2.6%	0.70
Specialist visits in ambulatory settings, per 1,000 beneficiaries										
Baseline	8,484	7,790				8,123	7,789			
Months 1–6	7,562	7,565	-3.2 (750)	< 1%	1.00	7,478	7,094	384 (612)	5.4%	0.53
Months 7–12	7,330	7,135	195 (835)	2.7%	0.82	7,074	6,573	501 (646)	7.6%	0.44
Months 1–12	7,450	7,358	92 (736)	1.2%	0.90	7,362	6,959	403 (602)	5.8%	0.50
Percentage of beneficiaries with any hospitalization stay in time period										
Baseline	32	31								
Months 1–6	14	15	-1.0 (3.3)	-7.1%	0.75					
Months 7–12	18	20	-2.2 (3.6)	-11%	0.54					
Months 1–12	27	30	-2.8 (4.1)	-9.5%	0.50					
Percentage of beneficiaries with any ED or observation visits in time period										
Baseline	53	57								
Months 1–6	27	29	-2.7 (4.0)	-9.2%	0.51					
Months 7–12	26	26	-0.81 (4.1)	-3.1%	0.84					

Table C.2 (*continued*)

	Results from using 2-year baseline period					Results from topcoding data at 98th percentile[a]				
	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value	Treatment group mean	Control group mean	Impact estimate (SE)	Percentage impact	*p*-value
Months 1–12	41	45	-3.4 (4.6)	-7.6%	0.46					
Sample sizes										
Number of beneficiaries										
Baseline	358	180				358	180			
Months 1–6	358	180				358	180			
Months 7–12	340	171				340	171			
Months 1–12	358	180				358	180			

Source: Mathematica's analysis of information from awardee's randomization file and Medicare claims and enrollment data from March 31, 2013, to February 28, 2017.

Note: Impact estimates are based on the regression-adjusted difference between the randomized treatment and control group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate).

[a] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline year and the follow-up year.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month; SE = standard error; SNF = skilled nursing facility.

Appendix D

Results of Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, program impacts for the University of California at San Francisco (UCSF) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. Drawing probabilistic conclusions requires external or prior evidence. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with similar background characteristics to UCSF. Probabilities were calculated using the results of a Bayesian regression that jointly modeled impacts on three core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for UCSF led to a Bayesian estimate of the program's impact on total Medicare expenditures of -1 percent (an estimated reduction of $20 per beneficiary per month) across the 12-month post-intervention period.

Table D.1 Comparison of frequentist and Bayesian impact estimates for UCSF in the first year after enrollment

	Impact estimate (95 percent interval)		Percentage impacts		
Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	-81 (-430, 268)	-20 (-176, 137)	< 1%	-5%	-1%
Hospital admissions	-46 (-186, 95)	-5.0 (-52, 42)	< 1%	-10%	-1%
ED visits	-63 (-260, 134)	-11 (-86, 65)	> -1%	-9%	-2%

Source: Mathematica's analysis of information from the awardee's finder file through September 20, 2017 and Medicare claims and enrollment data as of February 28, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

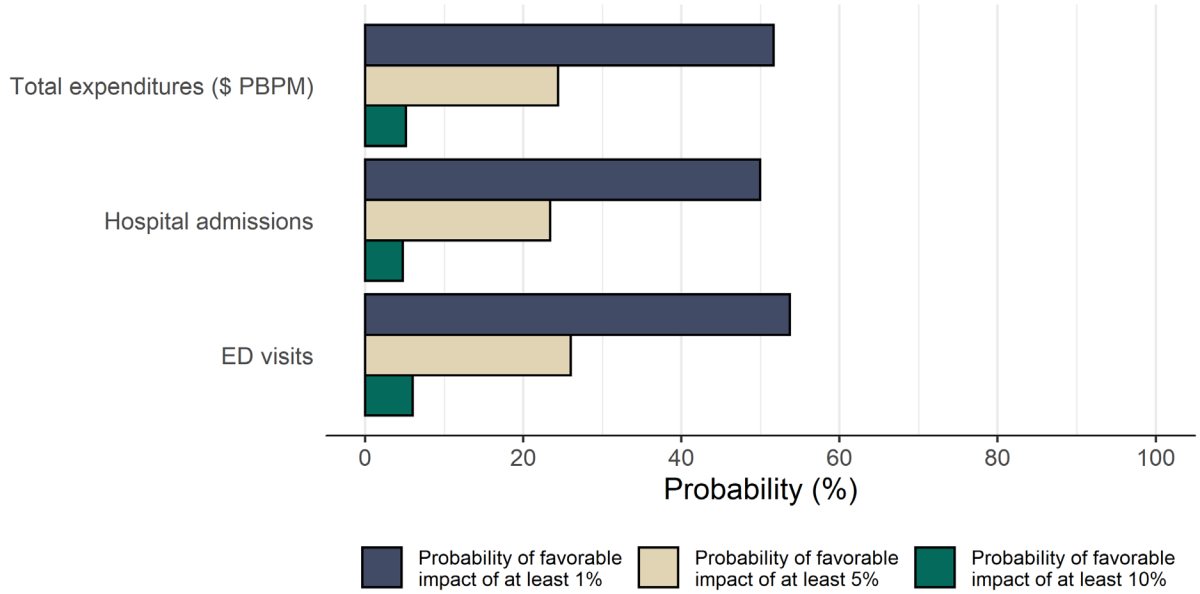
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results relied on a small sample and are therefore imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that UCSF achieved favorable impacts during the first year on three core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the UCSF program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from the awardee's finder file through September 20, 2017 and Medicare claims and enrollment data as of February 28, 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a moderate probability—in the range of 50 percent—that UCSF had a favorable impact of 1 percent or more on total Medicare expenditures, hospital admissions, and emergency department visits. Although suggestive, these probabilities are not large enough to indicate a substantial impact, especially in light of the small sample size and the correspondingly large standard errors associated with the impact estimates. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the UCSF program did not have a meaningful impact on total expenditures or service utilization.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Regents of the University of Michigan

September 2020

Boyd Gilman, Mynti Hossain, and Luke Horner

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

REGENTS OF THE UNIVERSITY OF MICHIGAN

The University of Michigan received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to expand the Michigan Surgical and Health Optimization Program (MSHOP). The University of Michigan began implementing the MSHOP at the University of Michigan Health System (UMHS) in 2011, three years before the award. The awardee used HCIA R2 funds to expand the MSHOP to 39 non-UMHS sites from September 2014 through July 2017. The primary goal of the program was to improve surgical outcomes for adults scheduled for major abdominal surgery. The target population consisted of individuals at participating surgical practices who were scheduled for a major abdominal surgery, were scored as high risk for poor surgical outcomes, and had at least one week between MSHOP enrollment and their planned surgery date. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The MSHOP represented an expansion of a prehabilitation program that the UMHS had operated within its system for two years.
- The program enrolled 3,051 adults at least one week before a planned major abdominal surgery, 1,200 of whom could be linked to the Medicare enrollment database. However, only 167 (14 percent) had a claim with an eligible procedure code from a participating surgeon.
- It was not possible to conduct a rigorous impact evaluation of the MSHOP because (1) a pilot program made it difficult to establish a basis of comparison before the sites implemented the program, (2) the risk assessment tool could not be replicated in claims data, and (3) the small number of enrollees in the analytic sample made it difficult to detect meaningful effects.

The MSHOP involved providers using a risk assessment tool at the point of referral or surgical consult for major abdominal surgery to assess patients' risk for postoperative complications, such as deep wound infection, myocardial infarction, pneumonia, and deep venous thrombosis. Providers invited patients who they assessed as being at high risk to participate in the MSHOP. The MSHOP taught participants healthy habits (such as those related to exercise and diet), giving them a standardized kit that contained supportive tools and materials. To encourage participants to engage with the intervention, the awardee designed an automated tracking system that reminded participants to record their activities and enabled them to track their progress online. The goals of the program were to reduce surgical complications, length of inpatient hospital stays after surgery, and payments for treatment of complications that occur in the hospital or after discharge.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The Regents of the University of Michigan implemented the MSHOP to assess participants' risk for postoperative complications after a major abdominal surgery (such as such as deep wound infection, myocardial infarction, pneumonia, and deep venous thrombosis) and, if medically appropriate, to engage them in a prehabilitation program that would potentially improve their surgical outcomes.
Major innovation	The MSHOP's key innovation was to educate presurgical patients at high risk for poor surgical outcomes to positively change their health behaviors by using tools and materials they received in a prehabilitation kit. In addition, the automated tracking system designed by the awardee was an innovative way to follow up with MSHOP participants to remind and encourage them to use the tools in the kit and to log their progress online or by telephone or text.
Program components	• Assessing risk for postsurgical complications • Optimizing healthy behaviors through educational materials and equipment • Monitoring health behavior changes through tracking and reporting
Target population	Adults at participating practices who met the following criteria: • Were scheduled for a major abdominal surgery • Were scored by practice staff as high risk for poor surgical outcomes • Had at least one week between MSHOP enrollment and surgery date
Total enrollment	The awardee enrolled 3,051 MSHOP participants, reaching 24 percent of its original enrollment goal. Of these, 1,200 could be linked to Medicare enrollment files, but only 167 of them had a claim for a qualifying major abdominal surgical procedure from a participating provider after the surgeon joined the HCIA R2-funded MSHOP.
Theory of change/theory of action	The UMHS hypothesized that the MSHOP's education, tools, and consistent reminders would motivate participants to engage in the prehabilitation activities of walking, completing breathing exercises, reducing or quitting smoking, improving nutrition, and reducing stress during the time between their enrollment in the program and their surgery. The University of Michigan believed that engaging participants in prehabilitation activities would lead to fewer surgical complications, reduce the length of inpatient hospital stays after surgery, and lower payments for treating complications that occur in the hospital or after discharge.
Award amount	$6,389,850
Effective launch date	September 15, 2014
Program settings	Surgical practices
Market area	A mix of urban, suburban, and rural
Target outcomes	The goals (based on results from a test of the MSHOP at the UMHS) were to reduce: • Surgical complications by 10 percent • The length of inpatient hospital stays by 2.3 days per case • The payments to hospitals for inpatient cost of care by $2,561 per case
Payment model	The awardee initially implemented a payment model in partnership with the BCBSM that paid surgeons and their care teams an incentive based on their level of engagement in the MSHOP. The awardee abandoned this payment approach after receiving feedback from surgeons that the payments were insufficient to incentivize participation. Instead, the awardee proposed using a FFS model, whereby the surgeon could bill the BCBSM for enrolling plan members in the MSHOP.

Table 1 (*continued*)

Program characteristics	Description
Sustainability plans	The UMHS sustained a version of the MSHOP, but non-UMHS practices did not. The awarded changed some of the MSHOP features based on lessons learned during the award. The awardee eliminated less effective program features (for example, the requirement to risk stratify patients and the use of automated technology to follow up with participants) and modified others to make the program more comprehensive and attractive to participants and providers (for example, implementing video visits to enroll patients into the MSHOP and expanding program services to include education and opioid and pain management).

FFS = fee-for-service; HCIA R2 = Round 2 of the Health Care Innovation Awards; BCBSM = Blue Cross Blue Shield of Michigan; MSHOP = Michigan Surgical and Health Optimization Program; UMHS = University of Michigan Health System.

It was not possible to conduct a rigorous impact evaluation of the MSHOP because of the way in which the awardee identified and recruited participants. As a result, this report describes only the enrollment, demographic, and health characteristics of Medicare participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis. Appendix A, Table A.1 describes the identification of the analytic sample.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	A rigorous impact evaluation of the MSHOP could not be conducted for three reasons. First, the evaluation team could not identify a credible comparison group; the team could not replicate in claims data the risk assessment tool used to identify participants. Second, the small number of Medicare FFS participants (167) identified in claims with a qualifying procedure code from a participating surgeon would have made it difficult to detect meaningful effects. Third, the team could not identify a clean pre-implementation period given that the UMHS implemented the MSHOP before HCIA R2. An analysis using all beneficiaries who met the claims-based eligibility criteria as the treatment group would have been unbiased, but was not feasible due to the low participation rate (7 percent). It is also likely that many of these nonparticipants did not meet the high-risk criteria and were therefore not actually eligible for the program.
Intervention group for descriptive analysis	The study included two separate but overlapping samples. The first included 1,200 participants in the awardee's finder file who could be linked to Medicare enrollment files; it represented 39 percent of the 3,051 total enrollees. The study used this sample to examine Medicare participants' enrollment characteristics that are available only in awardee-reported program data. The second sample included 795 Medicare FFS participants who could be linked to Medicare enrollment data *and* met the standard study inclusion criteria; it represented 26 percent of total MSHOP enrollment. The second sample excluded 271 participants who were enrolled in Medicare managed care, 88 who were not enrolled in both Parts A and B, 15 who did not have Medicare as a primary payer, and 31 with fewer than 90 days of claims history before enrollment. The study used this smaller sample to examine participants' characteristics that are only available in Medicare claims data, including chronic conditions, service use, and Medicare expenditures at the time of or during the year before enrollment.
Limitations	Due to the problems noted above, the analysis cannot be used to make inferences about the impact of this program on Medicare costs or other program outcomes.

FFS = fee for service; HCIA R2 = Round 2 of the Health Care Innovation Awards; MSHOP = Michigan Surgical and Health Optimization Program; UMHS = University of Michigan Health System.

PROGRAM DESIGN AND ADAPTATION

The MSHOP had three main components: (1) assessing risk for postsurgical complications, (2) optimizing healthy behaviors through educational materials and equipment, and (3) monitoring health behavior changes through tracking and reporting.[1]

Assessing risk for postsurgical complications

The MSHOP involved providers using a risk assessment tool on their mobile devices (such as smartphones and tablets) to assess participants' risk for poor surgical outcomes. The risk assessment tool was based on a predictive model developed by clinicians at the UMHS based on statewide outcomes data for patients who had a major abdominal surgery. The tool scored patients on a scale from 1 to 100, with 100 being the highest risk level. Clinicians considered patients scoring 50 or more to be at elevated risk for poor surgical outcomes and invited them to participate in the MSHOP. The awardee enrolled patients who scored less than 50 (or were never assessed) and provided them with the educational materials and equipment, but tracked them separately and did not consider them to be eligible for the MSHOP.

Optimizing healthy behaviors through educational materials and equipment

The MSHOP aimed to change the outcomes of care for surgical patients by encouraging them to engage in prehabilitation activities before surgery. Practice staff gave participants MSHOP packages that contained a pedometer and spirometer, as well as a compact disc and written materials encouraging them to walk more, perform breathing exercises, stop smoking, eat foods that promote health, and reduce their stress. For example, written materials included guidance on the types of foods to consume both before and after surgery to handle the stress of surgery, facilitate healing, and maintain muscle mass. Practice staff distributed the packages in person to patients who expressed to their provider at the point of referral or surgical consult that they wanted to participate in the MSHOP. The awardee allowed practice staff to choose whether they wanted to discuss the contents of the packages with participants or simply distribute the packages without much explanation. As a result, some participants received more in-person guidance on the design and benefits of the intervention than others—variation in program exposure that the study could not measure.

Monitoring health behavior changes through tracking and reporting

One of the requirements of the MSHOP was that participants had to submit data—via telephone, text, or visiting the MSHOP website—on how many daily steps they took and how many breathing exercises they performed using the spirometer. Practice staff and the MSHOP coordinating center relied on the tracking system to send automated notices to participants (and, if appropriate, their families) to remind them to submit their data. If participants did not respond to the automated communication within one week, practice staff or the MSHOP coordinating

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the MSHOP. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

center followed up with them by telephone or in person at their next medical appointment to ask them if they were having trouble submitting their data and help them address any challenges they faced.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee reported successfully implementing the MSHOP at the UMHS before the start of the cooperative agreement, giving it several years to address implementation challenges and refine mitigation strategies within the UMHS. However, the awardee struggled to expand the MSHOP beyond the UMHS for two reasons. First, lengthy processes to secure data use agreements and institutional review board approvals led to a slow enrollment pace of the non-UMHS hospitals' affiliated practices into the program. Second, implementation was delayed because the risk assessment tool needed for the program was not ready for use until May 2015—seven and a half months after the program launch. Due to existing hospital and health system copyrights, program leaders needed extra time to start their own company to disseminate the risk assessment tool as a non-UMHS product. In addition, the risk assessment tool required modifications.

Implications of program implementation for achieving program goals

- Previous experience implementing the MSHOP at UMHS-affiliated surgical practices made it difficult to identify a pre-period during which services were not already offered to eligible surgical patients.
- Expanding the eligibility criteria to increase enrollment could have led to fewer participants at high risk of postsurgical complications and, thus, fewer opportunities to improve outcomes.
- Delays implementing the risk assessment tool beyond the UMHS-affiliated practices, and the failure of many non-UMHS providers to use it, likely resulted in enrolling lower-risk patients less likely to benefit from the program.
- Challenges recruiting non-UMHS-affiliated hospitals and engaging their surgical practices in the program resulted in fewer participants than expected, particularly in the first two years of the program. This reduced the power of the study to detect statistically significant changes in outcomes, particularly in areas without the pilot test.

These challenges led to a lack of interest among non-UMHS providers to join the program and low engagement among those who did. Participating non-UMHS providers said they sometimes forgot to use the risk assessment tool when it was available to them and failed to recruit eligible patients into the MSHOP because they treated relatively few high-risk patients in their practices. The lack of engagement among non-UMHS providers, in turn, led to very low patient enrollment beyond the already-established UMHS network. To increase enrollment, the awardee expanded its eligibility criteria in the second program year to include additional abdominal surgeries, including major thoracic, urological, and vascular surgeries with an abdominal approach.[2]

[2] A full list of procedures eligible for inclusion under the HCAI R2-funded component of the MSHOP is available upon request.

DESCRIPTIVE ANALYSIS OF PARTICIPANTS' CHARACTERISTICS

Study sample

The descriptive analysis relied on two separate but overlapping samples. The first sample included all 1,200 participants in the awardee's finder file who the evaluation could link Medicare enrollment data; that sample represented 39 percent of the 3,051 total enrollees. The study used this sample to examine Medicare participants' enrollment characteristics that are available only in the awardee-reported program data. The second sample included only the 795 Medicare FFS participants who the evaluation could link to Medicare enrollment data *and* met the claims-based study inclusion criteria; it represented about one-quarter (26 percent) of total MSHOP enrollment. The second sample excluded 271 Medicare beneficiaries who were enrolled in Medicare managed care, 88 who were not enrolled in both Parts A and B, 15 who did not have Medicare as a primary payer, and 31 with fewer than 90 days of claims history before enrollment. The study used this smaller sample to examine participants' characteristics that are available only in Medicare claims data, including chronic conditions, service use, and Medicare expenditures at the time of or during the year before enrollment. (Appendix A, Table A.1 describes the identification of the Medicare FFS sample.)

Recruiting providers

By the end of the cooperative agreement, the awardee recruited 181 surgeons in the MSHOP who enrolled one or more patients in the study. Of these, 78 (43 percent) were affiliated with the UMHS and 103 (57 percent) were affiliated with non-UMHS hospitals. Although most of the UMHS providers had participated in the pilot MSHOP before the launch of the award, all non-UMHS-affiliated providers signed up for the program and began recruiting patients under the HCIA R2-funded expansion. Almost half (47 percent) of the non-UMHS surgeons joined the program and began recruiting patients in the second year of the program, and 40 percent joined in the third program year.

Recruiting and engaging participants

Although a slightly larger share of providers in the MSHOP were affiliated with non-UMHS hospitals, UMHS providers recruited and enrolled the overwhelming majority (82 percent) of the 1,200 patients who participated in the MSHOP and could be linked to Medicare enrollment data (Table 3). Non-UMHS-affiliated surgeons enrolled fewer than one in five (18 percent) of the Medicare beneficiaries who participated in the intervention. Nearly half (47 percent) of all Medicare participants did not enroll until the last year of the program.

According to program data from the awardee, only about one-quarter (27 percent) of the 1,200 Medicare participants received a risk assessment at enrollment. Of these, the median risk score was 50, indicating that, at enrollment, clinicians assessed half of all Medicare participants as not being at high risk for postsurgical complications and thus less likely to benefit from the

intervention. In addition, 30 percent of Medicare participants enrolled in the MSHOP with less than two weeks before their planned surgery date, and more than 60 percent had less than four weeks to engage in the prehabilitation activities.

Table 3. Program enrollment characteristics of all Medicare participants

	Participants (N = 1,200)
Affiliation of site where Medicare beneficiary was enrolled, %	
UMHS	82
Non-UMHS	18
Program year of enrollment, %	
First program year	26
Second program year	27
Third program year	47
Beneficiaries who received a risk assessment at enrollment, %	27
Risk score (0 to 100 scale)	
Mean	50
25th percentile	30
Median	50
75th percentile	71
Number of days between enrollment date and planned surgery date, %[a]	
Less than 2 weeks	30
2 to 4 weeks	31
4 to 8 weeks	22
More than 8 weeks	17

Source: Mathematica's analysis of information from awardee's finder file as of July 31, 2017.

Note: The evaluation counted the UMHS practices as one site because the health system consists of surgical lines or teams rather than individual practices.

[a] An additional 8 percent of Medicare enrollees had a missing planned surgery date.

UMHS = University of Michigan Health System.

Baseline characteristics of Medicare FFS participants

The mean age of the more restrictive sample of 795 Medicare FFS participants who the evaluation could link to Medicare enrollment data and met the study's claims-based inclusion criteria was 70, with 70 percent originally qualifying for Medicare based on age (Table 4). However, almost one in five Medicare FFS participants (18 percent) was younger than 65, nearly one-third (30 percent) became entitled to Medicare because of a disability, and one-fifth (20 percent) were dually eligible for Medicare and Medicaid.

Medicare FFS participants also had mean predicted expenditures more than twice the average of Medicare FFS beneficiaries nationally, as measured by the hierarchical condition category (HCC) score. Total average per beneficiary per month Medicare payment during the baseline year was $1,924, which was more than twice the 2014 national average of $816. In addition, 38

percent of the Medicare FFS participants had a hospital admission during the baseline year, and an annual rate of inpatient stays of 678 per 1,000 participants, compared with 282 stays per 1,000 Medicare FFS beneficiaries nationally. More than 40 percent had an emergency department (ED) visit, leading to an annual rate of ED visits of 956 per 1,000 participants. This profile of MSHOP participants is consistent with the awardee's intention to enroll patients at elevated health risk, but not too old to benefit from a program based on increased exercise and adopting other healthy behaviors.

Table 4. Baseline characteristics of Medicare FFS participants

Measure	Participants (N = 795)
Demographics	
Age at enrollment, years	70
Age group, %	
Younger than 65	22
65 to 74	53
75 to 84	23
85 and older	3
Female, %	55
White, %	90
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	67
Disability insurance benefits[a]	28
Medicare and Medicaid dual status, %	15
HCC score[b]	
Mean	2.1
25th percentile	0.9
Median	1.6
75th percentile	2.8
Service use and expenditures during year before enrollment[c]	
Total Medicare expenditures ($ PBPM)	1,924
Number of hospitalizations (per 1,000 beneficiaries)	678
Number of ED visits (per 1,000 beneficiaries)	956
Any hospitalizations, %	38
Any ED visits, %	43

Sources: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data as of July 31, 2017.

Notes: The descriptive analysis includes only Medicare FFS participants who met the standard claims-based study inclusion criteria, many of whom did not have a subsequent claim for an eligible surgical procedure. The evaluation defined the baseline period as the 12 months before the date the beneficiary enrolled in the program.

The statistics are weighted means, with participant weights proportional to the number of months during the study period that the participant was enrolled in Medicare.

[a] Includes participants with both a disability and ESRD.

Table 4 (*continued*)

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of 1 represents average expected expenditures. The evaluation calculated HCC scores by using the most recently available HCC algorithms.

ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Finally, according to awardee-reported program data, 335 (42 percent) of the 795 Medicare FFS participants in the descriptive analysis had their surgery suspended or cancelled after they enrolled in the MSHOP. Of the remaining 460 Medicare FFS participants, the evaluation could not find 293 (64 percent) in Medicare claims with a qualifying major abdominal procedure code from a participating surgeon after enrollment. Missing claims might have occurred if the procedure code was not reported on the claim, the surgery was performed on an outpatient basis, the participant withdrew from the program or died before surgery, or a professional claim with the participating surgeon's national identifier was not submitted. As a result, only 167 (36 percent) of the 460 Medicare FFS participants with a non-cancelled or postponed surgery met the claims-based eligibility criteria for the study.

Challenges measuring program impacts

Several factors made it difficult to evaluate the impact of the MSHOP. First, the program sought to enroll patients who were at high risk of post-surgical complications using a risk assessment tool that could not be applied to the claims data needed to identify a credible comparison group. Second, only 167 (36 percent) of the 460 Medicare FFS participants with a non-cancelled or postponed surgery had a claim indicating a major abdominal surgery performed by a participating provider, far too few to detect changes in outcomes of reasonable magnitude. Further, a 7 percent participation rate (defined as the 167 Medicare FFS participants who met the claims-based eligibility requirements divided by all Medicare FFS beneficiaries with a claim indicating a major abdominal surgery performed by a participating provider) meant that there would be a very low probability of detecting even a very large effect if the evaluation used an intent-to-treat approach to estimate impacts. Third, almost half of the surgeons participated in the MSHOP before the launch of the HCIA R2-funded program, making it difficult to determine the outcomes that participating surgeons would have achieved in the absence of the program. Most of the patients who received an eligible procedure from a UMHS-affiliated surgeon in the pre-implementation period would have been enrolled in the pilot program. The contamination of the pre-implementation group of eligible patients would bias the observed differences in outcomes between the pre- and post-implementation periods toward zero, relative to the ideal situation of no intervention during the baseline (pre-HCIA R2) period.

CONCLUSION

Evaluation challenges prevented assessing the impact of the MSHOP on the core outcomes. However, several implementation factors suggest that the impact of the program as implemented would have been small, even if the intervention had been effective. First, the delay in

implementing the risk assessment tool among non-UMHS providers, and their decision to rely on clinical impressions to identify and recruit high-risk patients, resulted in enrolling lower-risk patients into the program. Evidence of this is apparent in the small proportion of participants who received a risk assessment and the large proportion of those with an assessment who had a score below the cutoff used to identify high-risk patients. Second, expanding eligibility to include additional abdominal surgeries (for which the risk assessment tool had not been validated) might have also reduced the ability to identify and enroll patients at elevated risk for complications and thus weakened the intervention's effectiveness. Finally, the intervention itself provided a relatively light touch, offering only education materials, a pedometer, and a spirometer to promote and track healthy behaviors. The program enrolled most patients only two to four weeks before their planned surgery dates; it is difficult to change health behaviors and improve health outcomes from such behavioral changes in such a short period.

PROGRAM SUSTAINABILITY

After the award ended in August 2017, the awardee transferred ownership of the MSHOP to the Michigan Surgical Quality Collaborative (MSQC). The MSQC was a good candidate for sustaining the MSHOP because the program aligned with the MSQC's existing quality improvement work, and because the MSQC had helped recruit and enroll hospitals and affiliated practices into the MSHOP during the award period. The MSQC modified the sustained program by removing the requirement to risk stratify participants, removing the use of automated technology to follow-up with participants, initiating video visits to enroll patients into the MSHOP, and expanding program services to include education on opioid use and pain management. Only the UMHS sustained a version of the MSHOP and the non-UMHS practices did not.

UMHS abandoned the payment model it used during the cooperative agreement, under which Blue Cross Blue Shield of Michigan (BCBSM) provided incentive payments to participating practices based on their level of engagement with the program. The awardee abandoned this payment approach after receiving feedback from participating surgeons and care teams that the payments were insufficient to incentivize participation. The surgeons and care teams found the extra tasks required by the MSHOP to be burdensome and not fully covered by the payments, and low beneficiary enrollment into the program made it difficult to qualify for the maximum incentive payment.

University of Michigan's payment model

UMHS implemented a payment model in partnership with BCBSM that paid surgical providers an incentive based on their level of engagement in the MSHOP. Surgeons and their care teams earned points for completing the various steps in the program's workflow for each enrolled beneficiary and received an incentive payment that correlated with the number of points earned. Each surgeon was eligible for an annual incentive payment of up to $1,500, and the care team for an annual incentive payment from $2,000 to $4,000. However, the UMHS abandoned this approach and replaced it with a FFS model after surgical providers said the payments were insufficient to incentivize participation.

Based on this feedback, by the end of the third program year, the awardee decided to pursue a FFS payment approach with BCBSM instead. Initially, the awardee planned to bill G-codes for program services, but abandoned this approach after program leaders became concerned that the codes were inappropriately used for ineligible patients. Instead, the MSQC and BCBSM implemented a six-month pilot using a modifier code when billing for the procedure. Using the modifier qualified the physicians for additional compensation from BCBSM.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

The descriptive analysis relied on two separate but overlapping samples. The first sample included all 1,200 participants in the awardee's finder file who the evaluation could link to Medicare enrollment data; it represented 39 percent of the 3,051 total enrollees. The study used this sample to examine Medicare participants' enrollment characteristics based on awardee-reported program data. The second sample included only 795 Medicare fee-for-service (FFS) participants who the study could link to Medicare enrollment data *and* met the claims-based study inclusion criteria; it represented about one-quarter (26 percent) of total Michigan Surgical and Health Optimization Program (MSHOP) enrollment. The study used this smaller sample to examine participants' demographic, health, service use, and spending characteristics at the time of or during the year before enrollment.

Finally, according to awardee-reported program data, 335 (42 percent) of the 795 Medicare FFS participants in the descriptive analysis had their surgery suspended or cancelled after they enrolled in the MSHOP. The study could not find 293 (64 percent) of the remaining 460 Medicare FFS participants, in Medicare claims with a qualifying major abdominal procedure code from a participating surgeon after enrollment. Missing claims might have occurred if the procedure code was not reported on the claim, the surgery was performed on an outpatient basis, the participant withdrew from the program or died before surgery, or a professional claim with the participating surgeon's national identifier was not submitted. As a result, only 167 (36 percent) of the 460 Medicare FFS participants with a non-cancelled or postponed surgery met the claims-based eligibility criteria for the study. Any improvements in outcomes resulting from the prehabilitation intervention would have been limited to this small group, and would not have been detectable in program impact estimates.

Table A.1. Identifying the sample used in descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total Medicare beneficiaries in awardee's finder file		**1,200**
Did not meet study's standard claims-based inclusion criteria		
Not enrolled in both Part A and B	88	1,112
Enrolled in Medicare Advantage	271	841
Medicare not primary payer	15	826
Fewer than 90 days of claims history before enrollment	31	795
Final Medicare FFS beneficiaries in descriptive analysis		**795**
Surgery suspended or cancelled after enrollment	335	460
Unable to find claim with qualifying procedure from participating surgeon[a]	293	167
Medicare FFS beneficiaries who had a claim with an eligible procedure code from a participating surgeon after enrolling in the program		167

Sources: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data as of July 31, 2017.

Table A.1 (*continued*)

[a] Missing claims might have occurred if the procedure code was not reported on the claim, the surgery was performed on an outpatient basis, the participant withdrew from the program or died before surgery, or a professional claim with the participating surgeon's national identifier was not submitted.

FFS = fee-for-service.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Seattle Children's Hospital

September 2020

Brian Bruen, Victoria Peebbles, Marisa Morrison, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

SEATTLE CHILDREN'S HOSPITAL

Seattle Children's Hospital used the funding it received through Round 2 of the Health Care Innovation Awards (HCIA R2) to implement Pediatric Partners in Care (PPIC), a program focused on improving care coordination and reducing unnecessary or redundant services for children with disabilities enrolled in Medicaid and the Supplemental Security Income (SSI) program. The two main components of the program were (1) care management and coordination and (2) provider education. The awardee worked in close collaboration with four Medicaid managed care organizations (MCOs) that served patients in Washington State to implement the program and develop a framework for a sustainable delivery and payment model for these services. The program launched in February 2015, after an initial planning period. Table 1 summarizes key features of the program.

The awardee hypothesized that improvements in care coordination and management across hospital, primary care, and community settings would lead to better outcomes and reduce costs for children with complex care needs. The program's goals were to (1) improve health outcomes of children with disabilities; (2) reduce medical costs through reducing or eliminating unnecessary, redundant, or ineffective treatments; and (3) develop a scalable management model for outpatient care that optimized the existing care delivery infrastructure. Seattle Children's Hospital proposed a per beneficiary per month (PBPM) care management fee model, but it was still in the development phase at the end of the award period.

Important issues for understanding the evaluation

- Seattle Children's Hospital implemented PPIC, a program that aimed to improve care coordination and management across hospital, primary care, and community settings for children with medical complexity.
- The PPIC program involved close collaboration with Medicaid MCOs to develop the intervention services and payment model. The degree of collaboration was innovative, but created some implementation delays.
- This analysis relies on 516 PPIC enrollees who were Medicaid and SSI-eligible children and who resided in King and Snohomish counties in Washington State, and 955 children who met the same eligibility criteria as the treatment group, except that they lived in Pierce county.
- Findings should be interpreted with caution due to the small sample size and because there is likely selection bias since only 17 percent of beneficiaries who appeared to be eligibile for the program actually enrolled.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Seattle Children's Hospital's PPIC program aimed to (1) improve the health outcomes of children with disabilities who receive SSI and participate in Medicaid; (2) reduce medical costs for these children by eliminating unnecessary, redundant, and ineffective treatments; and (3) develop a scalable management model for outpatient care that optimized the existing care delivery infrastructure.
Major innovation	The PPIC program involved close collaboration with MCOs to develop the intervention services and develop a new payment model for those services.
Program components	• Care coordination and management • Provider education and training
Target population	The PPIC program focused on a population of about 3,000 Medicaid beneficiaries ages 18 and younger in King and Snohomish counties in Washington State who also received SSI, were enrolled in one of four Medicaid MCOs, and were identified as being at high risk for negative health outcomes.
Participating providers	PCPs affiliated with four Medicaid MCOs in Washington State
Total enrollment	The awardee enrolled 813 children (85 percent of the original enrollment goal) and recruited 34 primary care practices (170 percent of the original participation goal).
Level of engagement	PPIC sought to maintain engagement of patients and their caregivers through efforts to not oversaturate caregivers with contact, and by tailoring each participant's frequency of calls and types of home visits to that participant's needs. This study did not have information on level of engagement of participants.
Theory of change/ theory of action	Better access to care coordination and management will result in only necessary and effective treatment use, better quality of life, and lower costs. Enhanced navigation through social services and care coordination will address the medical and nonmedical needs of children and adolescents with disabilities.
Award amount	$5,561,620
Effective launch date	February 1, 2015
Program settings	Provider-based
Market area	Urban and suburban, King County (Seattle and suburbs) and Snohomish County (Everett and suburbs), Washington
Target outcomes	• Improve measures of care coordination by 10 percent for the majority of participants • Improve measures of a child's quality of life by 10 percent for half of the participants • Reduce the overall cost of care by 9.7 percent
Payment model	PBPM care management fee adjusted for measures of quality, use, and spending
Sustainability plans	Seattle Children's Hospital discontinued the PPIC by the end of the award, although program staff made efforts to sustain elements of the program through different initiatives.

MCO = managed care organization; PBPM = per beneficiary per month; PCP = primary care provider; PPIC = Pediatric Partners in Care; SSI = Supplemental Security Insurance.

While 813 beneficiaries participated in the program, this impact analysis included 516 of the participants who met the intervention's claims-based criteria and had sufficient Medicaid data for analysis, and a comparison group of 955 beneficiaries who met the program's eligibility criteria but lived in a neighboring county and thus were unable to participate. Table 2 summarizes the key features of the impact evaluation. Appendix A, Table A.1 describes the identification of the study sample.

Table 2. Key features of the program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group.
Intervention group for evaluation	Whie 813 benefciaries participated in the program, the analysis include 516 Medicaid- and SSI-enrolled children who received PPIC services from February 2015 to August 2017, excluding 99 children not served by the three MCOs that provided data, 136 children who did not appear to meet any of the high-risk eligibility criteria, and 62 children who did not meet additional criteria needed to enable credible matching and estimation of baseline service use.
Comparison group	The impact analysis compared outcomes among participants to those of a matched comparison group of 955 children with similar characteristics from a neighboring county.These 955 matched comparison beneficiaries were drawn from a larger sample of 2,299 children from Pierce County who met one or more of the PPIC eligibility criteria but were not eligible to participate due to living outside the two target counties.
Limitations	If participants differed from eligible nonparticipants in ways not captured in Medicare administrative files and claims, the impact estimates might be biased. The low participation rate (about 17 percent) would have made it difficult to identify impacts if measured over all eligible beneficiaries.

FFS = fee for service.

PROGRAM DESIGN AND ADAPTATION

The care management and coordination and provider education components were part of PPIC from the beginning, but each changed during the award. The collaboration with MCOs was ongoing throughout the award.[1]

Care management and coordination

The awardee created four care teams, each with a nurse care manager and a community coordinator. The teams engaged participants' families or caregivers and primary care providers (PCPs) in an initial assessment to identify barriers to and gaps in the children's care and to develop a care plan. Afterward, the nurse care managers talked regularly with families and PCPs to review the plan, monitor progress in implementing the plan, and assess the child's ongoing needs. The community care coordinators helped families connect with community resources and

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

navigate the health care and social services systems. The awardee sought to maintain engagement of patients and their caregivers (for example, their families) over time by not overwhelming them with contacts. Program staff tailored each participant's frequency of calls and types of home visits to that participant's needs.

Participants in the first two years of the program received services from PPIC continuously, as long as they remained enrolled. In the final year of the award, the awardee began to focus on short-term interventions, such as home visiting for asthma patients. In the third year, PPIC also graduated participants when the care manager identified no further needs. Graduates could reenroll if the family identified new needs, but no longer had regular contact with the care team.

Provider education

PPIC staff worked with PCPs that had multiple children enrolled in the program to identify opportunities for education to improve care. Engaging practices in educational endeavors was a minor focus initially, but positive feedback from PCPs led to expanded activity to identify strengths and needs among practices, starting at the end of the first year of the cooperative agreement. In the second year of the award, the program also began to offer training events for school nurses on topics such as asthma and feeding tubes.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Initially, the awardee aimed to coordinate care directly for 1,600 children identified as being at high risk for negative health outcomes. Enrollment did not start for several months into the program because of delayed receipt of data from MCOs to identify eligible children. Participation then ramped up slowly because caregivers had to be contacted and care plans developed, and some SSI-eligible children had to be enrolled in SSI before enrolling in PPIC. The awardee lowered its direct participant enrollment projection from 1,600 at the beginning of the award to 960 at the beginning of the second award year, in response to changes in reporting to define direct participants only as those for whom the caregiver completed the assessment process and opted into the program. The enrollment projection stayed the same thereafter.

Implications of program implementation for detecting impacts

- Changes in the enrollment process and target population make it challenging to measure effects of PPIC because of differing inclusion criteria over time and potential delays between enrollment and the start of care coordination.
- Using a comparison group of children from a different county might affect the results if there are unobserved differences in the populations, health care treatment norms, and costs between these geographic areas.
- The impact of the payment model could not be measured because it was not implemented by the end of the cooperative agreement.

Starting in the third quarter of the program, the awardee switched to an opt-out enrollment process so PPIC could enroll children before the care team contacted them. At the same time, the awardee also began enrolling all PPIC-eligible children who received care from a PCP engaged in the program. This change facilitated enrollment but added imprecision to the timing and measurement of the intervention because families and/or PCPs might not have fully engaged with the care team for an unknown—and likely varying—amount of time after enrollment. In Year 3, PPIC again revised the process to focus on enrolling specific populations (for example, children with asthma) for whom a more immediate effect might be possible before the program ended.

PPIC staff initially recruited PCPs that recognized and appreciated Seattle Children's Hospital's effort to fill gaps in care and improve care coordination. Providers engaged in the program in turn connected PPIC staff to other primary care practices that they knew could benefit from the program. This purposeful recruitment helped to secure participation from many practices, but likely introduced a selection bias because providers were not randomly chosen. The incentive to participate was lower for practices with existing care management programs, such as federally qualified health centers, or for practices with fewer high-risk patients.

ESTIMATING PROGRAM IMPACTS

Enrolling participants

PPIC sought to engage children and adolescents younger than 18 from King and Snohomish counties, enrolled in both Medicaid and the SSI program, and identified as having a high risk for negative health outcomes. The awardee's high-risk eligibility criteria was defined as having an inpatient hospital stay or two emergency department (ED) visits within the past six months, or having a Washington State Predictive Risk Intelligence System (PRISM) score greater than 1, indicating higher-than-average expected costs. The awardee reported enrolling 34 PCPs, exceeding its target of 20. At first, the awardee did not enroll beneficiaries until the care team completed an initial assessment and had a care plan in place. In the third quarter—part-way through the first year of the award—the awardee started the passive enrollment process described earlier. The awardee reported that it enrolled 813 direct participants between February 2015 and August 2017, about 85 percent of its final enrollment projection.

Study sample

While the program enrolled 813 participants, this analysis relied on 516 children from King and Snohomish counties identified as PPIC participants by the awardee and 955 comparison beneficiaries. The treatment group excluded some participants because the three health plans provided data that covered only 714 (88 percent) of the 813 children directly served by PPIC. Also, among these 714 children, 136 (19 percent) were excluded because they did not appear to meet the criteria for PPIC eligibility in the 12 months before their reported enrollment date based on the records provided for analysis, and an additional 62 children (9 percent) were excluded because they had insufficient baseline data. The analysis drew the 955 matched comparison

beneficiaries from a larger sample of 2,299 children from Pierce county who met the PPIC definition of high risk, based on records provided by the three health plans. Pierce county was chosen as the catchment area for the comparison group since since it borders King County and is the second most populous area in the state; in comparison, treatment group children resided in either King County (the state's most populous county) or neighboring Snohomish County (the state's third most populous county).

The awardee estimated that the target population was about 3,000 Medicaid and SSI-eligible children in King and Snohomish counties. As noted earlier in this report, data provided by the three health plans included 3,447 children from these counties who appeared to meet at least one of the eligibility criteria for PPIC. Of these children, 2,869 never enrolled in PPIC.

Characteristics of treatment and comparison group beneficiaries

Comparing treatment and comparison group characteristics at baseline confirmed that the two groups were balanced (Table 3). The average age of treatment and comparison group members in the baseline year was 7 to 8 years. Matching produced good balance between the treatment and comparison groups across a diverse set of health conditions. However, relatively small numbers of individuals in both groups had a specific condition or set of conditions, which limited the numbers of children from the comparison population who could be matched with the treatment group and added imprecision to the estimated effects because use and costs were highly variable and there were not enough observations to discern the most likely range of effects. Average total and pharmacy expenditures in the baseline period were noticeably higher for the treatment group than the comparison group in percentage terms, but the differences were acceptable when accounting for the variation in both outcomes. Appendix B provides the full balance results measured over 12 months before enrollment.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Measure	Treatment (N = 516)	Comparison (N = 955)
Demographics		
Age at enrollment, years	8.1	8.1
Male, %	60	61
Reported conditions, %		
Central nervous system condition	37	40
Cardiovascular disease	22	22
Developmental disability	21	20
Gastrointestinal condition	33	35
Metabolic condition	45	48
Psychiatric condition	38	37
Pulmonary condition	39	43
Renal condition	30	35
Skeletal condition	35	36

Table 3 (*continued*)

Measure	Treatment (N = 516)	Comparison (N = 955)
Measures of risk		
Mean CDPS score[a]	2.5	2.4
Mean PRISM score[b]	1.7	1.8
Service use and expenditures during year before enrollment		
Number of hospital admissions (per 1,000)	512	523
Number of outpatient ED visits (per 1,000)	1,761	1,791
Total Medicaid expenditures ($ PBPM)		
Mean	$2,883	$2,354
25th percentile	$202	$135
Median	$593	$387
75th percentile	$1,864	$1,556

Sources: Mathematica's analysis of information from the awardee's program enrollment data and claims data from three (of four) participating Medicaid managed care plans from January 2014 through February 2018.

Notes: The baseline year is defined as the 365 days before each beneficiary's enrollment date. All beneficiary characteristics were measured during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level, two-tailed test.

Appendix B presents full balance results.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

[b] PRISM is a clinical decision support tool developed by the state that generates predictive risk scores based on claims data.

CDPS = Chronic Illness and Disability Payment System, ED = emergency department; PBPM = per beneficiary per month; PRISM = Predictive Risk Intelligence System.

Analytic approach

The results in this analysis relied on a differences-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Assuming that external trends affect both groups similarly, a comparison group well matched on observable and unobservable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables will capture differences on unobserved variables as well. The primary outcomes are total spending, number of hospital admissions, and number of ED visits.

The awardee defined the pre-enrollment period as the year before each participant's enrollment date and the post-enrollment period as the two years after. The health plans that provided data reported the enrollment date for the treatment group. As noted earlier, the enrollment process varied over the life of the program; the enrollment date generally captured either the date that the

beneficiary had a care plan in place following initial engagement(s) with the care coordination team, or the date that the awardee contacted the child's caregivers preceding the first engagement with a care coordination team. Each comparison beneficiary received a randomly selected pseudo-enrollment date based on the months in which records indicated the individual met one or more of the eligibility criteria. Appendix A describes the statistical models and outcomes used to estimate the effects of the program.

IMPACT RESULTS

None of the estimated changes in outcomes for PPIC program participants relative to comparison group members on Medicaid spending and the use of inpatient and ED services were statistically significant (Table 4). The estimated difference implied that total expenditures decreased by 15 percent relative to the comparison group in the first year of enrollment, but increased by 9 percent relative to the comparison group in the second enrollment year. However, neither estimate was statistically significant, in part because of the small sample size and in part because the wide range of health conditions in the treatment group meant that service use and expenditures varied widely. When expenditure outliers were truncated at the 98th percentile, the the estimates implied that expenditures were somewhat higher for the treatment than control group, but were still not statistically significant (see Appendix C). In contrast, the estimated difference in pharmacy expenditures was sizeable and statistically significant. Pharmacy expenditures increased by an estimated 53 percent among participants relative to the comparison group in the second follow-up year. One possible explanation for the estimated increase in pharmacy spending is that care managers helped patients to either start new medicines to manage their conditions or to better adhere to existing drug regimens. Appendix C presents the full results of the difference-in-differences analyses, including sensitivity analyses that top-coded outliers for expenditures at the 98th percentile. Appendix D presents the results from the Bayesian analysis.

Table 4. Estimated impact of the PPIC program on selected outcomes

	First year after enrollment	Second year after enrollment
Total Medicaid expenditures ($ PBPM)		
Difference-in-differences estimate ($)	- $410	$252
Percentage	-15%	9%
p-value	0.45	0.72
Pharmacy expenditures ($ PBPM)		
Difference-in-differences estimate ($)	$45	$181**
Percentage	14%	53%
p-value	0.39	0.02
Number of hospitalizations, per 1,000 beneficiaries		
Difference-in-differences estimate (rate)	34	97
Percentage	11%	34%
p-value	0.71	0.36

Table 4 (*continued*)

	First year after enrollment	Second year after enrollment
Number of ED visits, per 1,000 beneficiaries		
Difference-in-differences estimate (rate)	58	51
Percentage	4%	5%
p-value	0.75	0.79
Sample size		
Treatment	516	402
Comparison	955	517

Sources: Mathematica's analysis of information from the awardee's program enrollment data and claims data from three (of four) participating Medicaid managed care plans from January 2014 through February 2018.

Note: Difference-in-differences estimates are based on the regression-adjusted difference between the randomized treatment and control group members. Percentages are then calculated as the regression estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the difference-in-differences estimate). Appendix C presents full regression results. Appendix D presents results from the Bayesian analysis.

* Significantly different from zero at the .10 level, two-tailed test.

** Significantly different from zero at the .05 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month; PPIC = Pediatric Partners in Care.

Consistent with the theory of action, program staff and clinicians who responded to surveys and interviews overwhelmingly agreed that the PPIC program produced positive impacts on medical and nonmedical needs of children and adolescents with disabilities. However, they noted that care coordination might not affect service use or the overall cost of care. For example, one care manager indicated that educating a child with cystic fibrosis about medications or disease process will not necessarily reduce the child's chances of going to the ED because there still might be instances when it is necessary. Another care manager noted that it is often appropriate for high-risk asthma patients to visit the ED when experiencing respiratory distress, and care coordination alone will not reduce these types of visits. Program leaders interviewed during the award period believed they had begun to see cost savings as a result of PPIC, especially related to more proactive care management (anticipating issues rather than reacting afterward) and reductions in redundant specialty care visits. However, estimated savings were not significant, and the analysis could not estimate the timing of care management or changes in specialty care visits using the available data.

Several other factors might explain why the PPIC program did not produce favorable measurable changes in total spending or inpatient and ED use. First, the program did not enroll many children until its second year, which meant that fewer than 330 participants had at least 18 months of follow-up data. Interview respondents noted that it might take longer than the length of the cooperative agreement to see an impact on the cost of care. Also, some patients did not start receiving care coordination services until well after their enrollment dates.

Second, the focus of the program changed from a broader care coordination model in the first two years of the award to a more targeted, condition-specific intervention in the final year of the award. Effects for beneficiaries enrolled earlier in the program would have reflected the broad

Main findings from the impact evaluation

- There were no meaningful estimated impacts on total Medicaid expenditures PBPM, hospitalizations, or ED use for participants in the first year or two years following PPIC enrollment.
- The small sample size and high level of variability in health conditions and service use in the treatment populations made it unlikely that even large effects of the program would be detected by this analysis.
- Qualitative findings from earlier reports suggested many positive impacts on quality of care and quality of life that the claims data available for this analysis cannot quantify.

care coordination, whereas effects for those enrolled in the final year might have reflected the transition to interventions targeted to specific populations, such as children with asthma. These more-targeted interventions might change service use and expenditures in a shorter period, but most individuals enrolled in the final year of the award could not be followed for more than a few months. If effects were not immediate, they might not have been observable in this analysis.

Third, the PPIC program might have achieved the goals of improving outcomes and developing a scalable model for outpatient care without reducing costs. Care coordination might in fact increase costs for some participants if they were not previously obtaining needed services, such as preventive medications and routine or follow-up visits with PCPs or specialists. Program staff and clinician survey respondents said the PPIC program had a positive impact on quality of care, quality of life, and patients' satisfaction, but the analysis could not quantify these quality measures for the comparison group using the available data, and those measures were not the core outcomes of this study.

Program leaders reported that their survey of families' experiences with care coordination showed improvement in the proportion of families that identified having a single care coordinator arrange their care, that the care coordinator helped their children to obtain needed community services, and that their children's care coordinator advocated for the needs of their children. Again, these measures were consistent with the theory of action that enhanced navigation through social services and care coordination will address the medical and nonmedical needs, but they were not quantifiable with claims data or the focus of this study. Frontline staff also noted that they had success in educating patient families on appropriate ED use and that families demonstrated a better understanding of using the ED versus urgent care, or called the care team if they had a concern, rather than immediately going to the ED. However, the models used in this analysis did not find statistically meaningful differences between the treatment and comparison groups in annual ED use per 1,000 patients (Table 4).

CONCLUSION

Overall, the program had no discernible effect on the primary outcomes of total expenditures or use of inpatient and ED services. The only statistically significant estimated change in outcomes was that children who participated in PPIC had higher prescription drug expenditures in the

second year after enrollment compared to children from the comparison group. This outcome might reflect changes in medication use attributable to the program's care management. Program staff and clinicians strongly believed that PPIC had many positive impacts on outcomes such as quality of care and patient satisfaction, but the study could not quantify those impacts with the data available for this analysis.

Limitations of evaluation

The analysis has several limitations. First, only 17 percent of the beneficiaries who met the program's claims-based eligibility criteria actually participated, based on provided data. Participants likely differed from eligible nonparticipants in ways that administrative and claims files could not capture; this means that the analysis likely suffers from selection bias and cannot be generalized to the larger eligible population. Second, the combination of small sample sizes and substantial variability in baseline service use and costs meant there was limited statistical power to detect effects. Third, the short follow-up period combined with the delayed receipt of care coordination would bias results toward no effect if changes in service use or expenditures resulting from care management took time to occur. Finally, no data were available to indicate how often or how long a patient received care coordination and the degree of coordination varied by patients' needs, adding imprecision to the definition of the intervention and uncertainty about how long it might take before effects on use or spending should occur, if at all. This variability in treatment across patients and the change in the focus of the program from broader care coordination to targeted interventions in the last year of the award also would make it difficult to identify program features that are most important for achieving favorable outcomes and to replicate the intervention in other settings.

PROGRAM SUSTAINABILITY

The awardee worked closely with Medicaid MCOs in Washington State to develop a new payment model for the care management and coordination services offered under PPIC. Under this model, MCOs would have received a PBPM fee for each Medicaid child enrolled in the MCO and in SSI. The awardee would adjust the fee over time based average spending; rates of hospitalization, readmission, and ED use; and performance on process measures. Major challenges to developing the payment model included (1) having limited claims data covering the program period, due to the slow ramp up of the program; (2) integrating separate data from four MCOs; and (3) developing a model for a relatively small population with a wide range of conditions.

In the final year of the award, agreements were in place with four of the five Medicaid MCOs operating in Washington State for a pilot year during which the awardee and the MCOs would exchange data to enable estimating financial impacts if the proposed fees were adopted. Seattle Children's Hospital discontinued the PPIC program at the end of the cooperative agreement. Although care management and provider education services ceased, the awardee used an extension of the award through May 2018 to analyze program data and continued negotiations with the MCOs pertaining to the service delivery payment model. All four partnerships with MCOs ceased by the end of the extension period: Three of the four MCOs chose not to continue funding the program and the fourth, which wanted to sustain the program, chose not to due to worries about attracting a disproportionate share of medically complex children. The awardee continued to work with this payer on a value-based contract with goals similar to those of the PPIC program, but targeting a different, healthier population.

PPIC's Proposed payment model

The awardee proposed a payment model in which Medicaid MCOs would pay Seattle Children's Hospital a PBPM care management fee for each child who is enrolled in both the MCO and the SSI program. Each year, the fee would be adjusted for performance based on:

- Average spending for the enrolled population compared to a pre-intervention baseline (35 percent of fee adjustment)
- Hospital, readmission, and ED service use rates compared to a pre-intervention baseline (40 percent of fee adjustment)
- A set of program-specific process measures (25 percent of fee adjustment)

The awardee planned to negotiate the payment amount with each MCO, but did not determine a value by the end of the award period.

The awardee reported that program staff tried to sustain elements of the program through different initiatives. For example, the awardee shared lessons learned to inform a new but similar initiative implemented by Seattle Children's Hospital, and attempted to engage the state Medicaid agency in discussions about PPIC and the problems it sought to address. However, as of July 2018, it had been unable to do so due to competing priorities in the state, such as the opioid crisis, behavioral health issues, and homelessness.

Appendix A

Description of analytic sample and modeling strategy

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays rely on a difference-in-differences approach with beneficiary fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or emergency department (ED) visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay or any ED visit during the baseline period. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy.

The treatment group included about 63 percent of total participants (Table A.1). To participate in the Pediatric Partners in Care (PPIC) program, a child had to have at least one inpatient visit or two or more ED visits or have a Washington State Predictive Risk Intelligence System (PRISM) score of 1 or higher, although program staff also enrolled children who might not have met these particular criteria but used health care services often or had high pharmaceutical spending. The awardee provided data for 714 children identified as PPIC enrollees and served by three health plans. Among this group, the analysis dropped 136 (19 percent) because claims and program data did not indicate that the participant met the standard high-risk criteria for PPIC eligibility (an inpatient stay, multiple ED visits, or PRISM score greater than > 1) in the 12 months before their reported enrollment dates. An additional 62 children (9 percent) were not reported as Medicaid-eligible on their program enrollment date or had fewer than 90 days of enrollment in the baseline period—additional requirements applied to this analysis to allow for credible matching and estimation of baseline service use.

The matched comparison group consisted of 955 children with similar health characteristics from Pierce County, located directly south of King County. The comparison group did not include non-enrolled children from King and Snohomish counties because they could have received services from providers that participated in the educational component of PPIC.

Table A.1. Identification of final sample for impact analysis for SCH

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through August 31, 2017		**813**
Participants not found in claims provided by the three Medicaid managed care plans[a]	99	714
Participants not reported as Medicaid-eligible on their program enrollment date or who had fewer than 90 days of enrollment in the baseline period [b]	62	652
Participants who did not meet the standard inclusion criteria (based on claims analysis)	136	516
Final analytic sample		**516**

Sources: Mathematica's analysis of information from the awardee's program enrollment data and claims data from three (of four) participating Medicaid managed care plans from January 2014 through February 2018.

[a] The claims data provided by the three Medicaid managed care plans included 714 individuals flagged as PPIC participants. A fourth participating Medicaid managed care plan did not provide data, so data were not available for any patients served exclusively by that plan. It is also possible that this analysis excludes any claims paid by this health plan for beneficiaries who switched to or from that plan during the analysis period. Patient identifiers did not allow tracking of individuals across health plans.

[b] The awardee did not apply these requirements to determine eligibility, but this analysis added these criteria to allow for credible matching and estimation of baseline service use.

PPIC = Pediatric Partners in Care.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. The table displays the weighted means of baseline characteristics for the 516 treatment beneficiaries and the 955 matched comparison beneficiaries used to estimate impacts. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable, which the study calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables include demographic characteristics (age and gender); health status as measured by the Chronic Illness and Disability Payment System (CDPS) score, Washington State Predictive Risk Intelligence System (PRISM) score, and chronic condition indicators; Medicaid expenditures in total and by type of service; and service use. The variables are measured over various specified intervals within the 12 months before enrollment in the intervention. For more details on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

The table also shows the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence of the statistical significance of the difference in the means. The equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the analysis conducted an omnibus test in which the null hypothesis was that the treatment and matched comparison groups balanced across all linear combinations of the covariates. The results assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

In general, these statistics and test results indicate that the treatment and matched comparison group aligned relatively well on these key characteristics. Percentage differences between the groups were relatively high in some cases, mostly when very few children had a particular health condition (for example, cancer). Standardized differences were acceptable (less than 0.10) or better for all of the characteristics.

Table B.1. Baseline characteristics of treatment and matched comparison groups for SCH

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age (years)	7.9 (0.21)	7.7 (0.16)	0.19 (0.29)	2.5	0.04	0.51	< 0.01
Male, %	60 (2.2)	60 (1.6)	0.32 (2.9)	< +/-1	0.01	0.91	< 0.01
Health status and diagnosis							
CDPS score[a]	2.5 (0.12)	2.3 (0.06)	0.16 (0.17)	6.5	0.06	0.33	< 0.01
PRISM risk score at enrollment	1.8 (0.05)	1.7 (0.03)	0.01 (0.08)	< +/-1	0.01	0.92	< 0.01
AIDS or other infectious disease, %	5.4 (1.00)	5.8 (0.56)	-0.42 (1.5)	-7.7	-0.02	0.77	< 0.01
Cancer, %	4.1 (0.87)	3.2 (0.51)	0.84 (1.2)	21	0.04	0.47	< 0.01
Cerebrovascular condition, %	4.3 (0.89)	3.3 (0.54)	1.0 (1.2)	23	0.05	0.39	< 0.01
Central nervous system condition, %	37 (2.1)	39 (1.5)	-1.2 (3.0)	-3.3	-0.03	0.68	< 0.01
Cardiovascular disease, %	23 (1.8)	22 (1.2)	0.84 (2.6)	3.7	0.02	0.75	< 0.01
Diabetes, %	2.7 (0.72)	3.2 (0.50)	-0.48 (1.1)	-18	-0.03	0.65	< 0.01
Developmental disability, %	21 (1.8)	19 (1.1)	1.4 (2.5)	6.8	0.03	0.57	< 0.01
Eye condition, %	3.9 (0.85)	2.6 (0.45)	1.3 (1.1)	33	0.07	0.24	< 0.01
Gastrointestinal condition, %	33 (2.1)	34 (1.4)	-0.94 (3.0)	-2.8	-0.02	0.75	< 0.01
Genital condition, %	3.3 (0.79)	2.5 (0.44)	0.84 (1.00)	25	0.05	0.40	< 0.01
Hematological condition, %	6.6 (1.1)	7.0 (0.72)	-0.42 (1.5)	-6.4	-0.02	0.79	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Metabolic condition, %	46 (2.2)	47 (1.6)	-1.1 (3.1)	-2.5	-0.02	0.71	< 0.01
Psychiatric condition, %	37 (2.1)	36 (1.6)	1.4 (3.0)	3.8	0.03	0.64	< 0.01
Pulmonary condition, %	39 (2.1)	41 (1.5)	-2.5 (3.1)	-6.3	-0.05	0.43	< 0.01
Renal condition, %	29 (2.0)	33 (1.3)	-4.7 (2.9)	-16	-0.10	0.11	0.01
Skeletal condition, %	34 (2.1)	34 (1.4)	0.81 (3.1)	2.3	0.02	0.79	< 0.01
Skin condition, %	8.5 (1.2)	8.9 (0.84)	-0.39 (1.8)	-4.5	-0.01	0.83	< 0.01
Substance abuse, %	1.2 (0.47)	1.3 (0.26)	-0.10 (0.67)	-8.3	-0.01	0.89	< 0.01
Medicaid expenditures							
Total inpatient spending, baseline year ($PBPM)	1,423 (480)	1,340 (182)	83 (578)	5.8	0.01	0.89	< 0.01
Total pharmacy spending, baseline year ($PBPM)	278 (35)	237 (20)	41 (45)	15	0.05	0.36	< 0.01
Total FFS spending, baseline year ($PBPM)	3,301 (573)	2,672 (210)	629 (693)	19	0.05	0.36	< 0.01
Service use							
Total hospitalizations	556 (53)	584 (59)	-28 (116)	-5.1	-0.02	0.81	< 0.01
Outpatient ED and observation visits	1,822 (104)	1,820 (67)	2.4 (168)	< +/-1	0.00	0.99	< 0.01
Propensity score	0.29 (0.01)	0.29 (0.00)	0.00 (0.01)	< +/-1	0.00	0.99	< 0.01
Number of beneficiaries	516	955					

Source: Mathematica's analysis of information from the awardee's program enrollment data and claims data from three (of four) participating Medicaid managed care plans from January 2014 through February 2018.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. p-values come from a weighted two-sample t-test; equivalence test p-values are the greater of the p-values for the two one-sided weighted t-tests of whether the

Table B.1 (*continued*)

true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending.

AIDS = acquired immunodeficiency syndrome; CDPS = Chronic Illness and Disability Payment System; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; PRISM = Predictive Risk Intelligence System (PRISM); SCH = Seattle Children's Hospital; SE = standard error.

Appendix C

Detailed results from impact estimates and sensitivity analyses

This page has been left blank for double-sided copying.

Table C.1 displays the impact estimates for the full study population, measured separately over intervention Years 1 and 2 and cumulatively over the full follow-up period. The analysis estimated the models over Medicaid expenditures, including total per beneficiary per month (PBPM) and total PBPM with high values top-coded at the 98th percentile to lessen the effects of high-cost outliers. Expenditures for prescription drugs were modeled separately because care management might be expected to produce changes in drug spending more quickly. Models were also estimated on service use measures including the number of hospitalizations and emergency department services used (per 1,000 beneficiaries). The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated change in select Medicaid FFS expenditures (dollars PBPM) and use measures associated with the SCH intervention during a 24-month follow-up period

	All beneficiaries				
	Treatment group mean	Comparison group mean	Estimated change in outcomes (SE)	Percentage change in outcomes[a]	*p*-value
Total Medicaid expenditures ($ PBPM)					
Baseline year	2,883	2,354			
Year 1	2,011	1,893	-410 (546)	-15%	0.45
Year 2	2,401	1,621	252 (706)	8.8%	0.72
Cumulative	2,109	1,817	-236 (603)	-8.3%	0.70
Total Medicaid expenditures, top-coded ($ PBPM)[b]					
Baseline year	2,249	2,044			
Year 1	1,817	1,744	-131 (239)	<1%	0.92
Year 2	1,954	1,426	324 (246)	16%	0.19
Cumulative	1,857	1,656	96 (263)	4.6%	0.69
Medicaid pharmacy expenditures ($ PBPM)					
Baseline year	275	247			
Year 1	372	300	45 (53)	14%	0.39
Year 2	489	281	181** (76)	53%	0.02
Cumulative	417	295	95* (54)	28%	0.08
Hospital stays, per 1,000 beneficiaries					
Baseline year	512	523			
Year 1	313	290	34 (93)	11%	0.71
Year 2	308	222	97 (106)	34%	0.36
Cumulative	310	263	58 (93)	19%	0.53
ED or observation visits, per 1,000 beneficiaries					
Baseline year	1,761	1,791			
Year 1	1,297	1,269	58 (180)	4.4%	0.75

Table C.1 (*continued*)

	All beneficiaries				
	Treatment group mean	Comparison group mean	Estimated change in outcomes (SE)	Percentage change in outcomes[a]	*p*-value
Year 2	1,056	1,035	51 (191)	4.6%	0.79
Cumulative	1,195	1,189	54 (168)	4.5%	0.75
Sample sizes					
Number of beneficiaries					
Baseline year	516	955			
Year 1	516	955			
Year 2	402	517			
Cumulative	516	955			

Sources: Mathematica's analysis of information from the awardee's program enrollment data and claims data from three (of four) participating Medicaid managed care plans from January 2014 through February 2018.

Note: Estimated change in outcomes for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The estimated change in outcomes for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage change in outcomes is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SCH = Seattle Children's Hospital; SE = standard error.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for Seattle Children's Hospital (SCH) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to SCH. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on three core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regression for SCH led to a Bayesian estimate of the program's impact on total Medicaid expenditures of between -5 and -6 percent (an estimated reduction of $113-$146 per beneficiary per month) in the first two years.

Table D.1. Comparison of frequentist and Bayesian impact estimates for SCH in the first two years after enrollment

Outcome	Follow-up period	Impact estimate (95 percent interval)		Percentage impacts		
		Frequentist	Bayesian	Prior	Frequentist	Bayesian
Total expenditures ($ PBPM)	Year 1	-410 (-1,481, 661)	-146 (-397, 100)	-8%	-15%	-6%
	Year 2	252 (-1,133, 1,636)	-113 (-339, 105)	-7%	9%	-5%
Hospital admissions	Year 1	34 (-147, 216)	-16 (-45, 13)	-8%	11%	-6%
	Year 2	97 (-111, 305)	-11 (-33, 11)	-7%	34%	-5%
ED visits	Year 1	58 (-296, 411)	-78 (-204, 47)	-8%	4%	-6%
	Year 2	51 (-324, 426)	-55 (-160, 45)	-7%	5%	-5%

Source: Mathematica's analysis of information from the awardee's program enrollment data and claims data from three (of four) participating Medicaid managed care plans from January 2014 through February 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

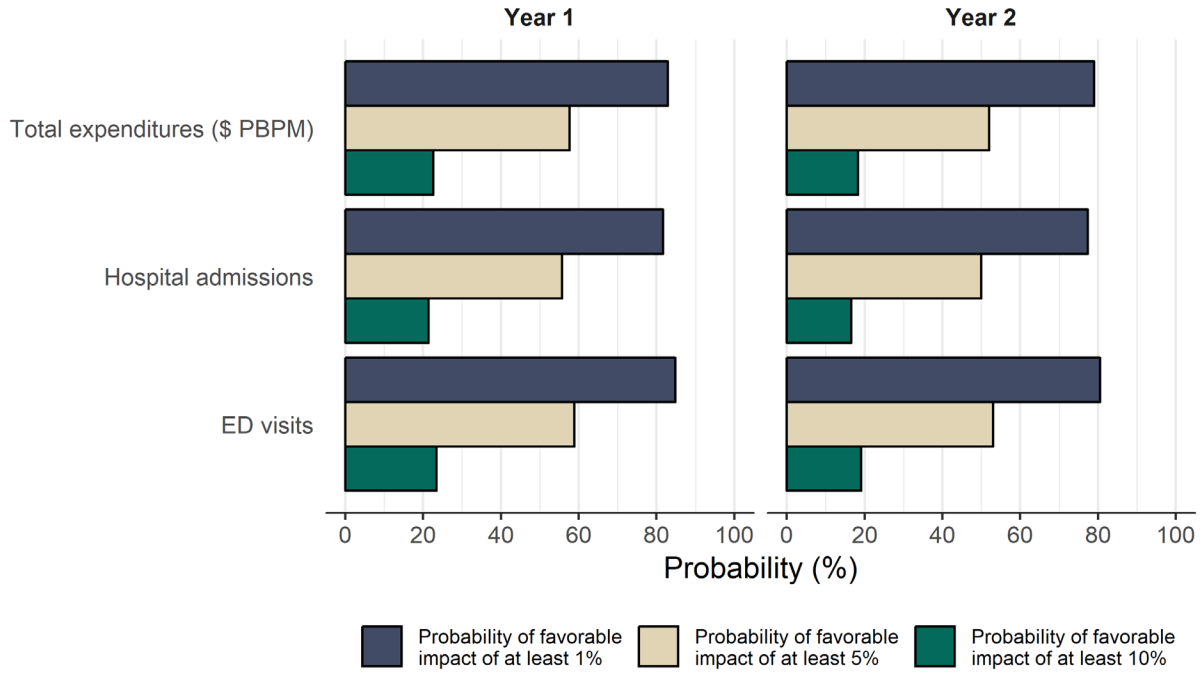
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results relied on a small sample and are therefore imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that SCH achieved favorable impacts during each of the first two years on three core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the SCH program had a favorable impact on key outcomes



Source: Mathematica's analysis of information from the awardee's program enrollment data and claims data from three (of four) participating Medicaid managed care plans from January 2014 through February 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a high probability—in the range of 80 percent—that SCH had a favorable impact of 1 percent or more on total Medicaid expenditures, hospital admissions, and emergency department visits, but the probability that the program has a sizeable impact of 10 percent or more on any of these outcomes was small—generally 20 percent or less. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the SCH program did not have a sizeable favorable impact.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Trustees of Columbia University in the City of New York

September 2020

Cordon Newhart, Kyle Emery, James Derzon, Brian Bruen, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

The Trustees of Columbia University in the City of New York received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the MySmileBuddy program. Under the intervention, community health workers (CHWs) worked with caregivers (parents or legal guardians) of young children to conduct risk assessments of the children's oral health, provide dental education, and develop strategies to prevent the progression of early childhood caries (ECC). The main innovation of the MySmileBuddy program was the use of a family-level, peer-counseling disease management approach (rather than traditional restorative dental procedures) to improve young children's oral health. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program used community health workers, supported by a tablet-based software suite, to conduct risk assessments of the children's oral health, provide dental education, and develop strategies to prevent the progression of ECC in enrolled children.
- Lack of support for the disease prevention model and structural incentives to conduct dental procedures in the referring pediatric dental delivery systems resulted in fewer referrals than anticipated.
- A rigorous impact analysis could not be conducted because (1) participants were selected into the program for reasons that cannot be observed (particularly the presence of early childhood caries) and (2) key outcomes related to oral health (such as the progression of dental disease) cannot be adequately captured in the claims.

Awardee leaders hypothesized that educating caregivers about ECC and oral health and engaging them in setting goals and planning to improve oral health care would change both caregivers' and children's behaviors. Caregivers would take steps to improve their children's oral health, such as encouraging and monitoring frequent and effective toothbrushing. The children's improved oral health behaviors would slow or stop the ECC disease process and reduce the need for restorative treatment and, in the most severe cases, costly surgery with additional risks from use of general anesthesia. The awardee's logic model assumed that avoiding these undesirable outcomes would reduce overall dental costs and improve children's oral health, which might also lead to better physical and psychological health.

Table 1. Program characteristics at a glance

Program characteristic	Description
Purpose	The program sought to improve young children's oral health by using family-level peer-counseling and behavioral risk-reduction strategies that supported using health education technology.
Major innovation	The MySmileBuddy program used a chronic disease management model and a tablet-based software suite within pediatric dentistry to improve the care experience and reduce costs.
Program components	• Engaging patients and families to recognize the value in preventing the progression of ECC and to implement strategies to prevent ECC • Health IT to conduct risk assessments, provide dental education, and develop strategies to prevent the progression of ECC in enrolled children
Target population	• The program sought to engage children ages 2 to 6 with ECC and no comorbidities whose caregivers spoke English or Spanish and were 18 years or older. • The program engaged up to two siblings younger than 6 of eligible children in the same household, regardless of whether they had caries if younger than the eligible child.
Participating providers	Columbia engaged five hospital-based PDDS clinics to identify and refer eligible families and provide standard care. Except for educating dental providers on the benefits of nonsurgical disease management for children's oral health care and encouraging them to refer participants to the MySmileBuddy program, direct engagement of providers was not a core component of Columbia's service delivery model.
Total enrollment	The awardee enrolled 1,207 participants, representing 62 percent of its original enrollment goal.
Level of engagement	Most participants attended scheduled encounters and participated actively in the MySmileBuddy program. The awardee reported making 5,401 contacts with 975 participating families. A few families withdrew from the program and some were lost to follow-up. Participants who completed the end-of-program survey conducted by the awardee reported high levels of satisfaction with the MySmileBuddy program.
Theory of change or theory of action	The awardee hypothesized that educating caregivers about ECC and engaging them in setting goals and planning to improve oral health would change caregivers' and their children's behaviors, leading to reduced ECC and potentially better physical and psychological health.
Award amount	$3,870,446
Effective launch date	The program began to operate in May 2015, after an eight-month planning period.
Program setting	• Recruitment conducted at PDDS clinics, day care centers, and community health fairs • Services delivered at participants' homes, in the community, or over the telephone
Market area	Urban, New York City (Bronx, Kings, New York, Queens, and Richmond counties)
Target outcomes	• Increased access to dental care • Development of caregiver-defined goals and action plans • Improvement in caregivers' use of preventive dental health behaviors with their children • Increased percentage of children who demonstrate no new cavities • Reduced dental costs of treating ECC
Payment model	The awardee proposed a PBPM Medicaid fee to dentists to help cover the costs of the MySmileBuddy software and CHWs who provide preventive dental services, but the model was not implemented.

Table 1 (*continued*)

Program characteristic	Description
Sustainability plans	In May 2018, the awardee received a five-year grant from the National Institutes of Health to continue the program and study the program's underlying "mechanisms of action." Columbia considered modifying how it identified participating children and the types of individuals it hired to serve as community health workers.

CHW = community health worker; ECC = early childhood caries; health IT = health information technology; PBPM = per beneficiary per month; PDDS = pediatric dental delivery systems.

It was not possible to conduct a rigorous impact evaluation of the MySmileBuddy program because of the way in which the awardee identified and recruited participants. As a result, this report describes only the demographic and health characteristics of Medicaid participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis.

Table 2. Key features of the descriptive analysis

Features	Description
Descriptive analysis	A rigorous impact evaluation of this program was not possible, primarily because key eligibility criteria (particularly evidence of ECC) could not be replicated in Medicaid claims data, and key outcome variables related to the progression of dental disease were not recorded in the Medicaid claims data. An analysis using all eligible children as the treatment group would have been unbiased but was not feasible due to the low participation rate among eligible families.
Intervention group for descriptive analysis	Of the 1,207 program participants, the descriptive analysis was limited to the 579 who had enough Medicaid data to construct reliable measures of Medicaid service use, met the program's eligibility requirements that could be replicated in the claims data, and enrolled in the MySmileBuddy program from May 11, 2015, to October 31, 2016. The analysis excluded an additional 442 beneficiaries who lacked valid Medicaid identifiers or did not have Medicaid as a primary payer, 91 beneficiaries without any baseline service use, 54 who did not meet the residency or age eligibility criteria, and 41 without dental services on the day of or the year before enrollment. Although having ECC or a sibling with ECC was an eligibility requirement, the analysis did not impose this condition because Medicaid claims did not reliably record an ECC diagnosis.
Limitations	Due to the problems noted above, the analysis cannot be used to make inferences about the impact of this program on Medicare costs or other program outcomes.

ECC = early childhood caries.

PROGRAM DESIGN AND ADAPTATION

The MySmileBuddy program service delivery model had two principal components: (1) participant and family engagement and (2) health information technology (health IT).[1]

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of MySmileBuddy program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Engaging participants and families

To engage and educate caregivers, CHWs provided information to caregivers about the MySmileBuddy program, confirmed their eligibility, and enrolled those who agreed to participate. The CHWs (1) used MySmileBuddy software to assess ECC risk for each eligible child; (2) educated caregivers about dental caries, including prevention strategies; (3) computed individualized risk scores for eligible children; (4) assisted caregivers in setting family goals; and (5) developed an action plan with the family to achieve those goals. CHWs followed up with families in person or by phone to assess progress toward goals and troubleshoot any problems in implementing the action plan. Meeting mode, frequency, and duration depended on families' needs and availability, but CHWs attempted to meet with the families in person at least once every three months to provide new toothbrushes and toothpaste.

Health IT

The MySmileBuddy program expanded its software to support CHWs and caregivers in a community setting by providing detailed and accurate preventive oral health information and templates for risk assessments and action plans for caregivers and children. The software also had an engaging participant interface for caregivers and children. An end-of-program survey found nearly all responding caregivers rated MySmileBuddy software easy to understand, visually appealing, and helpful for reaching their oral health goals.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee successfully implemented the MySmileBuddy program and CHWs and caregivers reported satisfaction with it. Most participants attended scheduled encounters and participated actively in the program. The awardee reported making 5,401 contacts with 975 participating families; CHWs met with each participant on average more than five times. Participants who completed the awardee's end-of-program survey reported high levels of satisfaction with the MySmileBuddy program, and caregivers reported that their children's oral health improved during the program.

Recruiting providers and screening and enrolling patients

The awardee faced several challenges expanding into community settings, all of which limited its ability to reach the enrollment target. First, the awardee had difficulty recruiting providers. Columbia University originally planned to partner with 12 pediatric dental delivery system (PDDS) clinics but found many PDDS clinics were unwilling to commit to participating in the intervention. The awardee found pediatric dental residency programs judged residents' performance based on the number of procedures they conducted, and residency program directors are judged by the dollars those procedures generate, so the financial and training incentives for PDDS staff did not align with disease management and prevention. In the end, the awardee partnered with five PDDS clinics that identified children who might benefit from the

MySmileBuddy program and connected interested caregivers with the awardee, which in turn distributed referrals to four community-based organizations that employed and supervised a total of 11 CHWs. The five PDDS clinics also provided dental care for children in the program.

Second, the awardee had difficulty getting dental residents to screen and refer patients into the program. Dental residents doing the screenings at the PDDS clinics had many demands on their time and often did not have time or forgot to talk to caregivers about the MySmileBuddy program, and fewer children than expected at the Head Start day care centers had caries. To overcome these barriers and increase referrals, in the second program year the awardee added screenings at a pediatric medical clinic, a mobile van, and community health fairs. To increase referrals from the PDDS sites, the awardee used several techniques to encourage dental residents and other PDDS staff to make referrals and to advertise the MySmileBuddy program directly to families. The awardee also hired a dentist to conduct additional screenings. Finally, the awardee expanded the program's eligibility criteria to include up to two siblings of the index child from the same household.

Despite these efforts, the awardee continued to have difficulty getting families to participate in the program. CHWs and program administrators reported many eligible families did not see the value of the MySmileBuddy program. CHWs and administrators felt caregivers frequently believed they provided their children with appropriate care by taking them to the dentist periodically for cleanings and cavity repair. The CHWs also noted that many families faced multiple challenges in their lives that took priority over the MySmileBuddy program. Adding to these challenges, the five PDDS clinics served large catchment areas with screening sites spread around the city, so families were often at a distance from where the CHWs were based and services delivered. CHWs and program administrators said travel times, program requirements, and the time commitment were frequent reasons that caregivers chose not to enroll, along with not wanting to participate in a research study or government-funded program and not understanding what the MySmileBuddy program involved. CHWs found that if dental residents explained the program and its value to caregivers and their eligible children, caregivers were much more receptive when the CHW called to enroll them. For families living far from the project, CHWs could travel up to an hour each way to meet with the eligible family.

The awardee intended to stop enrollment activities in May 2016, 15 months before the cooperative agreement ended, so that all participants could receive a minimum of 12 months of the intervention services. However, the awardee was far short of its enrollment target in May 2016 and thus extended enrollment through the end of October 2016. As a result, many participants received only 6 to 11 months of services, but the awardee felt 6 months were enough to change caregivers' behaviors in ways that would improve children's oral health, so the awardee did not expect reduced exposure to influence results.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Study sample

Although the awardee reported enrolling 1,207 participants, 219 of them were not included in the finder file for the evaluation. An additional 199 participants did not have a valid Medicaid ID and 24 did not have Medicaid as a primary payor in the enrollment data. In addition, 41 participants did not have a dental service on the day of or year before enrollment, and 54 did not meet the program's age and residency requirements (age 2 to 6 and lived in New York City). Finally, 91 participants had fewer than 90 days of Medicaid coverage before enrolling in the program, making it impossible to calculate their service use during the baseline period. (Appendix A, Table A.1 provides details on identifying the analytic sample.)

This analysis therefore includes the remaining 579 participants (48 percent) who enrolled in the MySmileBuddy program from May 11, 2015 to October 31, 2016, had sufficient Medicaid data to construct reliable measures of Medicaid service use at baseline, and met the program's eligibility requirements that the claims data could replicate. Although having ECC (or having a sibling with ECC) was an eligibility requirement for the program, the analysis did not use this criterion to define the analytic sample because Medicaid claims data do not reliably record ECC; only half of all participants had a diagnosis of ECC recorded in the claims data. In addition, only 237 (41 percent) of the 579 participants in the study sample had a dental visit during the year before or on the day of enrollment with a participating provider.

Characteristics of Medicaid participants

The overall health of the Medicaid participants was fairly typical for all Medicaid beneficiaries in this age group, with an average Chronic Illness and Disability Payment System score of 1.1, similar to the national average of 1.0 (Table 3). However, their dental service use was slightly higher than the Medicaid average. Although participants had to have a dental visit during the baseline year or on the date of enrollment for the analysis to include them, 64 percent had a preventive care visit the year before baseline, suggesting that many participants received dental care before enrollment. In comparison, 44 percent of Medicaid or Children's Health Insurance Program children in the state of New York had a preventive dental visit during 2016.[2] Participants rarely had a dental emergency department visit; the rate was only 16 per 1,000 beneficiaries during the baseline year. Finally, although the awardee recruited children with evidence of ECC, only 49 percent of participants had evidence of EEC according to Medicaid claims data, likely because Medicaid claims data do not reliably record ECC. Younger siblings of children with caries were also eligible to enroll, even if the younger sibling did not have caries.

[2] From the American Dental Association website, accessed on March 5, 2020, at https://www.ada.org/~/media/ADA/Science%20and%20Research/HPI/Files/HPIGraphic_0718_1.pdf?la=en

Table 3. Baseline characteristics of Medicaid participants

Characteristics	Medicaid participants (N = 579)
Demographics	
Age at enrollment, years	3.9
Age group, %	
Younger than 3 years, %	15
3 to 4 years, %	22
4 to 5 years, %	31
5 to 6 years, %	25
6 years or older, %	7.2
Female, %	52
White, %	2.2
Black, %	3.9
Hispanic or Latino, %	56
Asian, %	1.9
Other, %	0.70
Unknown, %	35
Original reason for Medicaid eligibility, %	
Disabled	4.4
Health status	
Had evidence of early childhood caries in baseline year, %	49
Mean CDPS score[a]	1.1
Medical health care service use during year before enrollment	
Any hospitalizations, %	3.5
Any outpatient ED visits, %	40
Oral health care service use during year before enrollment	
Any dental visit, %	100
Any preventive oral visit, %	94
Number of dental ED visits (per 1,000 beneficiaries)	16
Number of caries preventive services (per 1,000 beneficiaries)	1,613
Oral health care service use during year before enrollment	
Any preventive oral visit, %	64

Sources: Mathematica's analysis of information from awardee's finder file and Medicaid claims and enrollment data through October 2016.

Notes: The analysis defined the baseline year as the 365 days (12 months) before each beneficiary's enrollment date. It defined the pre-baseline year as the 12 months before the baseline year. It defined the enrollment date as the date of a participant's enrollment in MSB. The analysis measured all beneficiaries' characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; MSB = MySmileBuddy program.

Challenges of estimating program impacts

It was not possible to conduct a rigorous impact evaluation of the MySmileBuddy program for several reasons. First, there was a high likelihood that dentists referred children to the program, and their caregivers chose to participate after referral, for clinical or socioeconomic reasons not captured in Medicaid claims data, leading to selection bias. Second, although 1,207 children enrolled in the MySmileBuddy program, the key program eligibility criteria, having ECC, could not be used as a sample restriction because only half of participants had a diagnosis of ECC according to the claims data. Finally, an analysis on all eligible children who visited a participating PDDS clinic (which would eliminate the selection bias) was not possible because only 2 percent of the estimated 11,000 eligible children who visited a participating PDDS clinic actually enrolled in the program.

CONCLUSION

Overall, the MySmileBuddy program appeared to effectively engage participants and improve access to care. CHWs successfully engaged participants, meeting with them an average of 5.5 times during the program. Consequently, caregivers reported improvements in participants' oral health. The awardee's theory of action hypothesized that increases in the use of oral evaluation and preventive services for caries among participating children would change oral health behaviors, slow the progression of dental disease, and eventually lead to reduced costs. However, a rigorous impact evaluation of this awardee was not possible due to individuals qualifying for the program for reasons not observed in Medicaid claims data, the fact that the Medicaid claims data cannot capture key program outcomes related to the progression of dental disease, and the small size of the sample.

PROGRAM SUSTAINABILITY

Although Columbia University stopped operating the MySmileBuddy program by the end of its award in May 2018, the awardee received other funding to eventually resume the program. Specifically, the awardee received a five-year grant from the National Institutes of Health to study which program features contribute to or impede the program's success in reducing ECC. Under the program's new iteration, the awardee considered modifying how it identifies children to participate and the types of individuals it hires to serve as CHWs.

Columbia University also continued to pursue a payment model with Medicaid and other potential payers. The awardee abandoned its original plan to apply for a Medicaid state plan amendment, which would have allowed licensed health care professionals (in this case, dentists) to bill for preventive procedures delegated to nonlicensed health workers (in this case, CHWs). Instead, the awardee planned to approach Medicaid managed care and dental care organizations to add this function to their contracts.

The awardee also received an internal grant to explore how to expand the MySmileBuddy program to other social service programs that include an oral health component, such as Head Start and the Special Supplemental Nutrition Program for Women, Infants, and Children program, and other payers. To generate this interest, the awardee worked to determine how to assess whether the MySmileBuddy program generates savings, and how caries progression among MySmileBuddy participants compares to that of nonparticipating children with similar backgrounds.

Columbia University's proposed payment model

The awardee proposed paying dentists a per beneficiary per month Medicaid fee to help cover the costs of the MySmileBuddy software and CHWs who provided the preventive dental services. To meet state expectations for value-based payment approaches, the awardee identified several behavioral and clinical outcome criteria that dentists would have needed to meet to receive this payment.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

The program had 1,207 participants, but the roster for the evaluation did not include 219 of them, 199 did not have a valid Medicaid ID, and 24 did not have Medicaid as a primary payor on the enrollment data. In addition, 41 did not have a dental service on the day of or year before enrollment, and 54 did not meet the program's basic age and residency eligibility requirements. Finally, 91 had fewer than 90 days of enrollment in the baseline period, so baseline service use could not be analyzed. After imposing these restrictions, this analysis was limited to the 579 Medicaid beneficiaries who enrolled in the MySmileBuddy program from May 11, 2015, to October 31, 2016; who were ages 2 to 6 at the time of enrollment; lived in New York City; were enrolled in Medicaid; were without private insurance; and received a dental service (from any dentist, regardless of pediatric dental delivery system [PDDS] affiliation) in the baseline period. The final group of 579 children represented 73 percent of the 789 Medicaid beneficiaries in the MySmileBuddy program who had Medicaid identifying information (Table A.1).

A rigorous impact analysis was possible because so few beneficiaries would have remained in the treatment group after imposing all of the eligibility criteria observed in the Medicaid claims data. In particular, only 285 children met all of the program's eligibility criteria and had a diagnosis of ECC according to the claims data. Moreover, a design that used all eligible beneficiaries that visited participating providers could not be conducted because only 2% of beneficiaries that visited a participating provider enrolled in the program; conversely only 41% of beneficiaries that enrolled in the program had a claim for a visit to a participating provider on the day of or year before enrollment.

Table A.1. Identification of sample for descriptive analysis

	Number of participants excluded from analytic sample	Number of participants remaining in analytic sample
Total program participants		**1,207**
Not included in list of participants from awardee for the evaluation	219	988
Missing valid Medicaid ID	199	789
Was not a resident of New York City or ages 2 to 6 years at enrollment	54	735
Did not have a dental service on day of or year before enrollment	41	694
Did not meet standard claims-based inclusion criteria for study		
Not enrolled in Medicaid or enrolled in Medicaid but had private insurance on date of enrollment; or died before reported enrollment date	24	670
Had fewer than 90 days of Medicaid enrollment in baseline period	91	579
Final analytic sample for descriptive analysis		**579**

Sources: Mathematica's analysis of information from awardee's finder file and Medicaid claims and encounter data from May 11, 2015, to October 31, 2016. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: University Hospitals Cleveland Medical Center

September 2020

Lisa M. Lines, Lee-Lee Ellis, Allison Steiner, Justin Kirschner, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

UNIVERSITY HOSPITALS CLEVELAND MEDICAL CENTER

University Hospitals Cleveland Medical Center (UHCMC) used funding from its Round 2 Health Care Innovation Award (HCIA R2) to create the Learning Individual Needs and Coordinating Care (LINCC) program. The intervention provided care coordination and early palliative care to enhance the quality and experience of care while reducing its cost. The program's target population included adult Medicare and Medicaid beneficiaries receiving care at Seidman Cancer Center (and its community satellite locations) for the following complex cancers: (1) late-stage solid tumors or cancers with disease progression; (2) regionalized malignancies with complicating comorbidities; and (3) cancers complicated by other risk factors for poor outcomes and higher spending (for example, poor social support or low socioeconomic status). The program began operating in February 2015, six months after award. The intervention period covered under HCIA R2 ended in February 2018, after a six-month no-cost extension. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The program aimed to improve care quality and reduce costs for Medicare and Medicaid beneficiaries with complex cancers in Cleveland, Ohio, through care management.
- The program was new at UHCMC and became the standard of care, but patients could opt out of receiving services.
- Because it was not possible to use Medicare claims or clinical registry data to identify a comparison group that was similar to the intervention group at the time of enrollment, this study does not present impact estimates.

Awardee leaders hypothesized that, by providing cancer care that adheres to evidence-based guidelines and provides a more patient-centered approach, including early palliative care and proactive patient management with the support of a care coordinator, the program could (1) improve clinical outcomes and patients' satisfaction across the care continuum, (2) increase the quality of care (including coordination and evidence-based practice), and (3) decrease the total cost of care for patients with complex cancers.

The proposed payment model had two elements: (1) a per member per month (PMPM) care coordination fee and (2) shared savings between provider and payer. The PMPM fee was in addition to any available fee-for-service (FFS) payments, and included all services in the intervention, including coordinating care; educating, engaging, and assessing patients; and managing care to improve compliance with care plans. Providers were to split shared savings according to a negotiated rate, subject to predetermined quality standards. However, payment model discussions with payers did not yield any agreements.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	UHCMC provided care management and coordination to patients with complex cancers to improve quality of care, improve patients' satisfaction, reduce total cost of care, and demonstrate the feasibility and sustainability of a new payment model.
Major innovation	The program aimed to achieve higher quality and better experience of care through adherence to evidence-based guidelines, including early palliative care and care coordination.
Program components	• The program offered care management from a nurse care coordinator who helped the participant establish a plan of care, served as the participant's point of contact and advocate, facilitated patient and family engagement and education, linked the participant and family to resources, and ensured that outpatient care was well coordinated. • Participants received early and ongoing access to expert-level palliative care. • The program also offered health IT, including a routine biopsychosocial assessment administered on a tablet computer.
Target population	The awardee sought to engage adults receiving complex cancer care at Seidman Cancer Center and at least two of its community satellite clinics. Eligible patients included complex cancer patients, defined as patients with late-stage (3 and 4) solid tumors or disease progression, regionalized malignancies with complicating comorbidities, and other risk factors for poor outcomes and higher spending.
Total enrollment	The awardee enrolled 1,340 patients from February 2015 through September 30, 2017, representing 75 percent of its original enrollment goal.
Theory of change or theory of action	Nurse care coordinators identify adults receiving care for complex cancers and assess their physical, emotional, and spiritual needs. The coordinators communicate these needs to other members of the disease team. Nurse care coordinators work with patients and the primary oncology care team to develop a plan of care and connect patients with resources. Nurse care coordinators provide an extra layer of support that helps patients adhere to their plan of care; improve self-efficacy; and manage their physical, emotional, and spiritual needs. Better management of patients' needs results in better quality of care, participants' satisfaction, and appropriate service use, which in turn results in improved health outcomes across the continuum of complex cancer care and lower health care costs.
Award amount	$4,675,383
Effective launch date	February 23, 2015
Program settings	Cleveland, Ohio-based clinics associated with UHCMC
Market area	Cleveland, Ohio
Target outcomes	• Maintain or improve quality of care compared to that measured in 2013 baseline data (when available) and comparable peer group • Improve the participant-reported experience of care for a cohort of complex cancer patients by 5 percent from 2013 over experience of a comparable peer group • Improve the efficiency of health care delivery by reducing total cost of care for a cohort of complex cancer patients by 8 percent from 2013 costs • Demonstrate feasibility and sustainability of an innovative, asymmetrical, shared savings payment model to support enhanced service delivery • Decrease avoidable ED visits, hospitalizations, and 30-day hospital readmissions

Table 1 (*continued*)

Program characteristics	Description
Payment model	UHCMC's payment model consisted of PBPM payments of $160 for program services in addition to traditional FFS payments. Clinical services otherwise reimbursable by payers were not covered by the PBPM payment.
Sustainability plans	UHCMC planned to sustain several aspects of the program, including expert-level palliative care, regular biopsychosocial assessments, and enhancements to the electronic medical records to document advance care planning and goals of care. It also planned to sustain nurse care coordinators, but was working out how to finance this position and best use the nurse care coordinators' time.

ED = emergency department; FFS = fee-for-service; IT = information technology; PBPM = per beneficiary per month; UHCMC = University Hospitals Cleveland Medical Center.

It was not possible to conduct a rigorous impact evaluation of the LINCC program because of the way in which it identified and recruited participants into the program. As a result, this report describes only the demographic and health characteristics of Medicare FFS participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	It was not possible to conduct a rigorous impact evaluation of LINCC because of how the awardee identified and recruited beneficiaries into the program. A comparison group of beneficiaries similar to the intervention participants could not be identified using health care claims because hospital staff used clinical judgment to recruit participants. A comparison group selected to match participants on observed characteristics had a much lower mortality rate than the participants to whom they were matched (28 versus 53 percent). As a result, this report is limited to describing the demographic and health characteristics of Medicare FFS participants before they enrolled in the program.
Intervention group for descriptive analysis	The intervention group for the evaluation was based on the 488 participants (among the total enrollment of 1,340) who were enrolled in Medicare FFS for at least three months before and after enrollment into the program. The 852 excluded beneficiaries included 260 who were not Medicare beneficiaries and 583 who were not in Medicare FFS for at least three months before and three months after enrollment. Nine other beneficiaries were excluded because of missing data or an enrollment date preceding their first diagnosis date in claims data.
Limitations	Due to the problems noted above, no inferences can be made about the impact of this program on Medicare costs or other program outcomes.

FFS = fee-for-service.

PROGRAM DESIGN AND ADAPTATION

The LINCC program service delivery model had three main components: (1) care management from a nurse care coordinator, (2) early and ongoing access to expert-level palliative care, and (3) health information technology (health IT).[1]

Care management

LINCC nurse care coordinators worked with disease teams (doctors and nurses who specialize in treating a specific cancer) to manage participants' care. The nurse care coordinators identified participants' physical and emotional needs; connected participants to clinical and nonclinical resources that reflect those needs (for example, social work, spiritual care, and pain management); helped participants navigate their appointments and tests; served as a resource when participants had questions; helped participants establish goals for their care; and helped participants make more informed decisions about their care. Nurse care coordinators also engaged and educated patients and their families. In addition, they worked with patients to develop advance directives and promote adherence to patient-centered plans of care, which included patients' goals, future appointments, and current medications. The awardee determined that the best caseload for the nurse care coordinator was about 100 patients and that the exact caseload should vary with the patients' disease type and acuity.

Palliative care

Two palliative care providers delivered early and ongoing palliative care to better manage pain and other symptoms and address other domains of palliative care. Although palliative care already existed at the center, the LINCC program increased access to palliative care by encouraging patients to meet with palliative care providers earlier in their treatment and introducing palliative care as symptom management rather than end-of-life care. During the award, the LINCC palliative care providers began providing palliative care in the inpatient setting to improve the continuity of care across the inpatient and outpatient settings.

Health IT

To help identify and assess participants' needs, the LINCC program staff asked participants to complete a routine biopsychosocial patient assessment, which staff administered on a tablet computer. The assessment included questions from several validated tools designed to evaluate patients' physical needs and symptoms, emotional state, and social well-being. The nurse care coordinators and palliative care providers used a patient's responses to focus their services on the patient's needs. They also encouraged the oncologists at Seidman Cancer Center to review patients' responses.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The LINCC program was a new program implemented in a large academic hospital system. During the first nine months of the award, UHCMC successfully recruited and brought on board key personnel, developed an electronic patient assessment tool, and worked with the IT team to enhance the medical record to promote better care coordination. UHCMC successfully provided services, including coordinating care, identifying participants' needs, linking participants to resources, helping participants establish goals for their care, helping participants navigate their appointments, and increasing access to palliative care. Program staff reported successfully engaging participants and, eventually, engaging more providers as well.

Implications of program implementation for achieving program goals

- During the first program year, staff encountered challenges integrating program services into existing clinical workflows. This led to delays in implementing the program and reduced the likelihood of detecting impacts among early enrollees.
- Later in the program, nurse care coordinators and providers began using clinical judgment to identify participants who might benefit most from program services. Because this clinical judgment could not be replicated in comparison group data, it introduced a bias in any impact estimates attempted.

The LINCC program had several implementation challenges. For example, staff turnover prevented the LINCC program from fully expanding to its two largest community-based satellite clinics as planned. In addition, program staff reported difficulty integrating the program into the existing clinical workflow and physical infrastructure, especially at a large academic hospital with many providers and clinic locations. Clinical integration improved because LINCC staff were flexible about where and when to meet with patients to minimize disruptions. LINCC staff also provided outreach and education to providers about the purpose and value of the program. Nevertheless, at the end of the third program year, UHCMC still reported that "changing culture and securing a place for the services in the overall model of care" was its most significant challenge. Finally, program staff used qualitative and quantitative data throughout the award to refine their enrollment processes, initially enrolling as many qualifying participants as possible, to later focusing on participants who could most benefit from care coordination and palliative care.

Enrolling and engaging participants

The awardee identified individuals who met the program's eligibility criteria in two ways. Most often, the care coordination team reviewed reports generated by an internal data management system that contained information on patients' insurance and disease stage. If the report did not include the disease stage, the nurse care coordinator would confirm the stage with the disease team to confirm the patient's eligibility. The care coordinator would then contact the patient to

introduce the program. Alternatively, oncologists at Seidman Cancer Center could refer patients. If an oncologist believed a patient met the eligibility criteria and could benefit from the program, the oncologist could reach out to the LINCC nurse care coordinator and ask the coordinator to meet with the patient.

Nurse care coordinators reported that, during the third program year, they focused on recruiting patients they believed would benefit most from the intervention. Their focus evolved from enrolling as many patients as possible to taking a more active and targeted approach. The nurse care coordinators used chart reviews, physicians' notes, frequency of telephone calls between patients and the care team, and face-to-face interactions with the oncology care team to determine patients who would benefit the most from the program. The nurse care coordinators prioritized patients with characteristics associated with poor outcomes (for example, triple negative breast cancer), poor social support, and need for active symptom management. Program enrollment became more subjective over time because it relied on staff's clinical judgment and assessment of social factors.

The nurse care coordinators reported that the program successfully engaged participants by providing them an additional layer of support. However, the palliative care providers noted that it was difficult to engage asymptomatic patients and their providers. Nevertheless, by the end of the third program year, most participants (97 percent of participants enrolled in program year 3 and 70 percent cumulatively) had received at least one palliative care consult.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Characteristics of Medicare FFS participants in the LINCC program

The awardee enrolled 1,340 participants, but only 488 beneficiaries (36 percent) of the awardee's list of participants were enrolled in Medicare FFS for at least three months before and after enrollment into the program (Appendix A, Table A.1).[2] Among this subset of participants, the average age was 72 (Table 3). The mean hierarchical condition category (HCC) risk score was 3.6, indicating that UHCMC enrolled a Medicare FFS population predicted to be 3 to 4 times more costly in the upcoming year than an average Medicare FFS beneficiary. At the time of their initial diagnoses in the Ohio Cancer Incidence Surveillance System (OCISS) cancer registry, about half of beneficiaries had distant metastasis, and 13 percent were unstaged. Cancer stage at the time of entrance into the program cannot be determined. The most common cancer sites were lung, head and neck, breast, and colorectal. However, a third of beneficiaries had a disparate set of cancer sites categorized as other. During the 12-month period before entering the program, 59

[2] The evaluation limited the descriptive analysis to Medicare FFS participants who had three months of enrollment in Medicare FFS in both the baseline and program periods to produce more stable estimates of baseline characteristics.

percent of beneficiaries had a hospitalization and 50 percent had an emergency department visit. The average PBPM total Medicare expenditure was $3,606.

Table 3. Baseline characteristics of Medicare FFS beneficiaries

Measure	Medicare FFS participants (N = 488)
Demographics	
Age at enrollment, years	72
Age group, %	
Younger than 65	13
65 to 75	49
75 to 84	29
85 and older	9.2
Male, %	46
White, %	76
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	80
Disability insurance benefits or ESRD	20
Medicare/Medicaid dual status, %	
Dual eligible	13
HCC score[a]	
Mean	3.6
25th percentile	2.0
Median	3.3
75th percentile	4.6
Cancer stage at time of initial diagnosis, %[b]	
Local	14
Regional	25
Distant	49
Unstaged	13
Cancer site, % [b]	
Lung	18
Head and neck	14
Breast	11
Colorectal	11
Prostate	5
Lymphoma, leukemia, myeloma	3
Skin	3
Other	34
Service use and expenditures during the year before enrollment	
Any hospitalizations, %	59
Any outpatient ED visits, %	50
Beneficiaries with a readmission, %	10
Total Medicare expenditures ($ PBPM)	3,606
Number of outpatient ED/OBS visits (per 1,000)	937
Number of hospital admissions (per 1,000)	1,072

Table 3 (*continued*)

Source: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data from February 2015 through February 2018, as of October 15, 2019.

Notes: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the participant first met with a nurse care coordinator. All beneficiary characteristics were measured during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

[b] Percents do not sum to 100 due to rounding.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; LINCC = Learning Individual Needs and Coordinating Care; OBS = observation; PBPM = per beneficiary per month.

Challenges of estimating program impacts

It was not possible to conduct a rigorous impact evaluation of the LINCC program because of the way in which the awardee identified and recruited participants. Limitations in the data prevented matching on important variables likely associated with outcomes, such as unobservable risk factors for poor outcomes and higher spending levels, disease progression, and social determinants of health. Although the study had information on stage of cancer at diagnosis, the program focused on patients with complex cancers, including people whose disease had progressed to later stages of cancer. Because the cancer registry did not contain information on disease progression after diagnosis, the study was unable to use this critical characteristic when identifying a comparison group. As an indication of the challenges in identifying an appropriate comparison group, after applying rigorous methods to selecting a potential comparison group, a descriptive analysis showed that 28 percent of the potential comparison group died, whereas 53 percent of the treatment group died during the follow-up period. This suggests that matching on available characteristics could not account for differences between the two groups on unobservable risk factors. Further, program enrollment relied on staff's clinical judgment and assessment of social factors. For these reasons, this report does not present impact estimates.

CONCLUSION

UHCMC aimed to improve care quality and reduce costs for Medicare and Medicaid beneficiaries with complex cancers in Cleveland, Ohio. Its LINCC care management program included coordinating care, identifying participants' needs, linking participants to resources, helping participants establish goals for their care, helping participants make and keep their appointments, and increasing access to palliative care. Program staff reported successfully engaging participants and, eventually, engaging more providers as well. By the end of the third program year, most participants (97 percent of participants enrolled in program year 3 and 70 percent cumulatively) had received at least one palliative care consult. However, the LINCC program had several implementation challenges, such as staff turnover, which prevented the LINCC program from fully expanding to its two largest community-based satellite clinics. In

addition, program staff reported difficulty integrating the program into the existing clinical workflow and physical infrastructure, especially at a large academic hospital with many providers and clinic locations. Given how the awardee identified and recruited participants, it was not possible to conduct a rigorous impact evaluation of this program. A group of comparison beneficiaries similar to the intervention participants could not be identified in health care claims because hospital staff used clinical judgment to recruit participants.

PROGRAM SUSTAINABILITY

Anticipating insufficient funding to sustain its full program, UHCMC stopped enrolling new participants in the LINCC program six months before its award ended in February 2018, but continued some program services for existing participants with a mix of internal and external resources. The awardee integrated some workflows and tools into Seidman Cancer Center operations. With internal funding, it also sustained the expanded palliative care services at the main campus (and limited services at its satellite clinics), the streamlined patient assessment at all locations, and the electronic health record feature on goals of care. Funding from the Centers for Medicare & Medicaid Services (CMS)-sponsored Oncology Care Model (OCM) supported the salaries of nurse care coordinators who served in the same capacity as the LINCC nurse care coordinators at the main clinics and some satellite clinics. CMS launched OCM in July 2016. It involves monthly enhanced payments for oncology services and episode-based chemotherapy payments for high quality care.

UHCMC's proposed payment model

UHCMC proposed paying for LINCC through CMS's OCM payment approach, which combines enhanced capitated monthly payments with FFS payments. Capitated payments of $160 PBPM would cover clinical services that are otherwise not reimbursable by payers and would be supplemented by traditional FFS payments. Participating providers are eligible for semiannual performance-based payments (including one- and two-sided risk arrangements).

UHCMC also proposed a coordination fee and shared savings arrangement through its accountable care organization, in which commercial and public payers would cover the cost of the program for patients not eligible for OCM.

UHCMC was not able to use OCM funding for other aspects of its program, such as expanding nurse care coordinators to all satellite clinics and the spiritual and pharmacy intervention services. UHCMC had originally engaged commercial payers whose payments might have covered these aspects of the program. However, the payers ultimately declined to participate in the LINCC program because it would have required substantial redesign of their actuarial systems. In addition, although the awardee believed the accountable care organization was the most promising venue for financially supporting the LINCC program, UHCMC had decided to delay negotiations with the accountable care organization because of modest reticence to amending existing ACO arrangements with payors. Despite these setbacks, the awardee reported having continued strong organizational support for integrating select program services and pursuing value-based models of care.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants		**1,340**
Not enrolled in Medicare or could not be identified in Medicare enrollment files	260	1,080
Not enrolled in Medicare FFS or observable during 3 months in baseline and follow-up periods	583	497
Enrolled before date of initial cancer diagnosis or missing data	9	488
Final analytic sample		**488**

Sources: Mathematica's analysis of information from the awardee's program encounter database from September 1, 2014, through August 31, 2017, and Medicare claims and enrollment data from September 2013 through February 2018, as of October 2019.

FFS = fee-for-service; LINCC = Learning Individual Needs and Coordinating Care.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: University of Kansas Health System

September 2020

Brant Morefield, Vladislav Slanchev, Colene Byrne, Bob Schmitz, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

UNIVERSITY OF KANSAS HEALTH SYSTEM

The University of Kansas Health System, a medical center based in Kansas City, Kansas, received a cooperative agreement under Round 2 of the Health Care Innovation Awards to implement the Kansas Heart and Stroke Collaborative across the state of Kansas. The Collaborative grew over time and, by the end of the award, included 54 rural hospitals, 12 emergency medical service agencies, a tertiary care hospital, 19 rural physician practices, and two federally qualified health centers. The program sought to improve outcomes for patients with heart disease or stroke and, later in the program, sepsis or trauma; and to reduce the cost of care. Table 1 summarizes the key characteristics of the program.

Important issues for understanding the evaluation

- This impact evaluation focused on the first of the program's three arms to be implemented: the heart attack and stroke protocols (also known as the acute care phase of the program)
- The acute care arm was a provider-level intervention, based on implementing a standardized inpatient protocol. The program passively enrolled patients who met the eligibility criteria when they presented at the CAH and would not have known they benefitted from the intervention.
- This impact analysis relied on 920 Medicare FFS beneficiaries who presented at one of the 11 critical access hospitals in 14 rural counties in Kansas with stroke or heart attack symptoms and who had at least six months of program exposure.
- The comparison group included 2,247 Medicare FFS beneficiaries who were patients at 23 critical access hospitals in Nebraska, and thus ineligible to participate in the intervention.

Awardee leaders hypothesized that evidence-based protocols, provider education, telemedicine, transitional care management (TCM), and chronic care management (CCM) through health coaching and patient and family engagement would collectively (1) produce measurable improvements in rural Kansans' heart health and post-stroke survival and (2) drive significant reductions in total cost of care related to heart disease and stroke. Through an acute care intervention arm, the Collaborative supported providers in rural hospitals with clinical protocols, training, and tele-consultation designed to improve acute care for all patients who presented with time-sensitive heart attack or stroke symptoms, and later in the project, with sepsis or trauma. Through one ambulatory arm, the Collaborative provided short-term TCM for heart attack and stroke patients after discharge. In another ambulatory care arm, the Collaborative provided long-term CCM for patients who had suffered or were at risk of heart attack or stroke. Both ambulatory arms sought to engage Medicare and Medicaid beneficiaries and dually eligible beneficiaries.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The Collaborative program sought to improve outcomes for patients with heart disease, stroke, and, later in the program, sepsis and trauma; and reduce the cost of care.
Major innovation	The Collaborative program was innovative because it built trust relationships and program legitimacy and developed a long-term strategy for quality improvement in rural areas with a focus on time-sensitive, high-impact diagnoses.
Program components	• An acute care arm implemented standardized protocols for responding to suspected heart attacks and strokes in inpatient settings and provided help in treating these conditions from remote emergency or critical care specialists by using a telehealth solution. • Two ambulatory care arms included providing TCM after hospital discharge and CCM services to patients at risk for heart attack or stroke. • Health IT included emergency telehealth to support acute care arm and patient registries, health information exchange, population health dashboard, and data analytics to support the ambulatory care arms.
Target population	The target population for the acute care arm included residents of 14 rural counties (later extended to all counties) in Kansas who were hospitalized with or had symptoms of heart attack or stroke. The acute care arm included all patients, regardless of payer. The target population for the ambulatory arms was patients who had or were at risk of having heart attack or stroke and included Medicare and Medicaid beneficiaries and dually eligible beneficiaries only.
Participating providers	The program engaged rural hospitals, primary care providers, community health clinics, emergency medical services, a rural tertiary care hospital, and an academic medical center. The impact analysis relied on 11 CAHs and primary care providers that served patients from the 14 original counties.
Total enrollment	The awardee enrolled and served 7,334 participants across all three arms, nearly 100 percent of its original goal. The program directly enrolled participants for the TCM and CCM arms through consent forms and they received the services. The acute care arm participants enrolled indirectly by presenting at the rural hospital with diagnoses and/or symptoms of heart attack or stroke, and later sepsis and trauma, that were appropriate for following the Collaborative's clinical protocols. These indirectly enrolled participants included all eligible patients, not just those for whom the protocols were followed. There was likely overlap among patients across the three arms, but no way to determine this for the study.
Level of engagement	Based on interviews, staff and clinician surveys, and self-monitoring measures, the program trained clinicians at all of the CAHS in using the protocols. Over time, clinicians' acceptance and adoption of the acute care protocols increased, reinforced by regular onsite meetings with CAH staff to review their performance. TCM and CCM services expanded in the second year of the program because of improved analytics to identify eligible patients, support by primary care providers to provide these services, use of telehealth to reach more patients, and reimbursement incentives through the MSSP, so it is likely that many of the patients in the study received one or both of these services.
Theory of change or theory of action	The awardee hypothesized that evidence-based protocols, provider education, telemedicine, TCM and CCM services through health coaching, and patient and family engagement would collectively (1) produce measurable improvements in heart health and post-stroke survival and (2) drive significant reductions in total cost of care related to heart disease and stroke.
Award amount	$12,523,441
Effective launch date	March 1, 2015
Program settings	Rural hospitals and primary care practices

Table 1 (*continued*)

Program characteristics	Description
Market area	Rural counties in Kansas
Target outcomes	• 20 percent reduction in heart attack and stroke • Reduction in 30-day mortality following AMI • Clinical heart attack and stroke metrics, such as time to tests and administration of therapy • Reduction in all-cause unplanned readmission and inpatient days after readmission • Increase in discharged-alive rate for heart attack, coronary artery bypass graft, or percutaneous coronary intervention patients • Reduced rates of transfers from CAHs to inpatient hospitals • Reduced total cost of care for participants • Reduced rates of ED visits • Improvement in medication adherence
Payment model	The awardee created new Medicare FFS payments and shared savings under the MSSP. The awardee designed and analyzed data for CMS on a transformational rural health payment model with global budgeting, but did not implement the model under the award.
Sustainability plans	After the award ended, the awardee continued the program by using (1) revenue from CCM and TCM services furnished to FFS Medicare beneficiaries, (2) shared savings received through the MSSP, and (3) University of Kansas funding for specific staff and services. The awardee formed a new organization to operate as an ACO participating in the MSSP. The ACO supported the TCM and CCM services as well as ongoing performance improvement activities.

ACO = accountable care organization; AMI = acute myocardial infarction; CAH = critical access hospital; CCM = chronic care management; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; IT = information technology; MSSP = Medicare Shared Savings Plan; TCM = transitional care management.

The impact analysis presented in this report included 920 Medicare fee-for-service (FFS) beneficiaries who presented at one of the 11 critical access hospitals (CAHs) in 14 rural counties in Kansas with stroke or heart attack symptoms and who had at least six months of program exposure and met the other claims-based study inclusion criteria. The study identified a propensity score matched-comparison group from Medicare enrollment and claims data. The comparison group included 2,247 Medicare FFS beneficiaries with similar demographic and health characteristics who similarly presented at one of 23 CAHs with stroke or heart attack symptoms in neighboring Nebraska, where the program was not implemented. Table 2 summarizes the key features of the evaluation.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a post-implementation, cross-sectional analysis of the treatment population and a matched group of comparison beneficiaries. The estimates measured impacts as the difference in outcomes between treatment and comparison beneficiaries during the post-implementation period only. Models were run separately for beneficiaries who presented with stroke and heart attack and controlled for observed characteristics.
Intervention group for evaluation	The treatment group included 920 Medicare FFS beneficiaries. All Medicare FFS beneficiaries presenting at the ED with a stroke or heart attack diagnosis at any of 11 participating CAHs, and having no prior ED visit, observation stay, or inpatient admission with a similar diagnosis in the prior 12 months, were eligible for participation and included in the analytic sample. Medicare FFS status required Parts A and B enrollment for 90 days before and including the anchor event.
Comparison group	The impact analysis compared outcomes among participants to those of a comparison group of 2,247 Medicare FFS beneficiaries who were patients at 23 CAHs in Nebraska and thus ineligible to participate in the program. All comparison group beneficiaries also met the same study inclusion criteria applied to the treatment group.
Limitations	If participants differed from eligible nonparticipants in ways not captured in Medicare administrative files and claims, the impact estimates might be biased. Differences in average cost-based reimbursements to treatment and comparison CAHs could also have affected estimated changes in inpatient expenditures between the treatment and comparison populations.

CAH = critical access hospital; ED = emergency department; FFS = fee for service.

PROGRAM DESIGN AND ADAPTATION

The Collaborative sponsored three arms: (1) implementing acute care heart attack and stroke protocols, (2) the TCM arm, and (3) the CCM arm. The Collaborative referred to the first arm as the acute care arm and the latter two as the ambulatory care arms. Health information technology (IT) was a key feature of both the acute care and ambulatory arms and is discussed separately later.[1] The impact evaluation focused on beneficiaries who were eligible for the acute care arm. Receipt of TCM and CCM under the ambulatory care arms might also have affected outcomes, but the analysis was unable to measure the independent effects of these services on outcomes.

Acute care arm

For the acute care arm, clinicians at the University of Kansas Medical Center and Schools of Medicine, Nursing, and Health Professions developed evidence-based acute care protocols for detecting and treating stroke, heart attack, and later sepsis and trauma. The Collaborative then provided education and training to staff of CAHs, other rural hospitals, and emergency medical systems on the clinical practice protocols. The training recipients were primarily physicians, nurse practitioners, and emergency medical technicians. The protocols ensured that patients presenting at rural hospitals (typically for emergency care) receive appropriate, timely treatment preventing further morbidity and mortality and improving outcomes. Use of the protocols by clinicians was voluntary.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

In addition, rural hospital clinicians could receive help in treating these conditions from remote emergency or critical care specialists by using Avera's eCare eEmergency telehealth solution. This solution enabled clinicians to immediately connect with a board-certified emergency medicine physician and critical care nurse at the Avera eEmergency hub. The Avera specialists helped the rural hospital staff follow the protocols and document care delivery processes, freeing the rural hospital staff to work directly with the patient. Use of this service was left to the discretion of the treating clinician. The Collaborative also developed a protocol to standardize and improve the reliability and quality of information exchanged when patients transferred between rural hospitals and other providers.

Ambulatory care arms

Although the impact evaluation focused on the acute care arm, patients who presented at the CAHs with heart attack or stroke symptoms could have also received TCM or CCM services. The ambulatory care arms provided TCM services for 30 days following a hospital discharge to patients who agreed to enroll in the program. During the TCM period, the transitional care managers, who were advanced practice nurses, completed a home visit and made usually six to eight follow-up telephone calls with patients to monitor their health, ensure they understood and followed their medication regimen, and facilitate follow-up visits with their primary care physician and other specialists. The transitional care managers also worked with local hospital discharge planners to ensure patients had a comprehensive and integrated discharge plan.

The Collaborative also provided TCM for patients for 30 days following discharge from the hospital. In addition, it provided CCM for (1) patients who transitioned out of the TCM services; and (2) primary care outpatients who were at risk for heart attack or stroke, were referred directly by their primary care providers, or were identified by the Collaborative's population health information tools (for example, have hypertension diagnosis). A health coach (usually a registered nurse) delivered the CCM services, which included a home visit and telephone calls to assess the participant's physical and psychosocial well-being. TCM and CCM patients recovering from a heart attack or heart failure also received a formal resiliency training program from health coaches to help them and their family members better understand the disease and develop skills to manage symptoms and cope with functional limitations.

Health IT

Health IT was an important component of the acute and ambulatory arms of the program. The rural hospitals in the acute care arm could remotely connect with board-certified emergency medicine specialists at the Avera eEmergency hub for support in following the protocols and documenting care delivery processes. For CCM, eHealth coaching enabled trained coaches to provide CCM services using telehealth technology to reach more patients and reduce the amount of travel required by the health coaches.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The Collaborative met its goals and delivered services for all three arms in the initial catchment area, as well as across the state. Program leaders and staff reported that the awardee enrolled participants early enough in the program so that participants would receive enough exposure to services to expect improvements in clinical care. Participating clinicians and nonclinician staff reported that the program had successfully engaged patients and had a positive effect on care delivery.

Implications of program implementation for detecting impacts

- Most clinicians at the participating CAHs received training in the acute care protocols. However, the awardee did not require clinicians to use the acute care protocols and the degree to which they were used when beneficiaries presented with stroke or heart attack symptoms was not known.
- Process measures indicated that the 11 CAHs used the protocols, but their use varied within and across the CAHs. Some physicians resisted the protocols, especially early in the program. The measures also showed that there was some slippage in use over time.
- Although some patients in the acute care arm treatment group might have received services through the ambulatory arms (which could have improved their outcomes), it was not possible to identify and control for the other services received in the impact evaluation.
- Collaborative staff successfully implemented the program after they provided support and encouragement to the providers while holding them accountable for better outcomes.

Although it implemented the overall program successfully, the Collaborative experienced challenges getting rural hospital physicians to buy into the program and use the acute care protocols. The Collaborative's executive director and medical director spent time with clinicians, such as cardiologists, reviewing the protocols, explaining the best practices, and gaining buy-in for using the protocols. Program staff also provided support and encouragement to participating clinicians while holding them accountable for improvements in patients' outcomes. They recognized and celebrated large and small successes and offered resources when things did not go as well as hoped—for example, when providers fell back into old habits and did not follow the protocols. They also held the local sites responsible for their activities and for meeting performance benchmarks.

ESTIMATING PROGRAM IMPACTS

Enrolling participants

The acute care arm of the program was a provider-level intervention, based largely on implementing a standardized inpatient protocol for heart attack and stroke patients. The program passively enrolled patients when they presented at the CAH with diagnoses and/or symptoms of heart attack or stroke. The awardee expected them to receive services based on the protocols

implemented under the acute care arm. Patients would not have been aware that they benefitted from the intervention, nor did they have the opportunity to decline these services. As a result, the program considered all patients with an eligible condition regardless of payer to be indirect participants. No information was available on which indirect participants were treated with the acute care protocols or with assistance from Avera's eCare eEmergency telehealth solution. In addition, some acute care arm patients could have enrolled in the ambulatory arms of the program, but this information was also not available and the impact analyses could not control for this.

Study sample

The impact study relied on 920 Medicare FFS beneficiaries who presented in the emergency department (ED) at any of 11 participating CAHs, with or without a subsequent admission, for stroke or heart attack diagnosis from March 2015 to August 2017, six months before the award ended in February 2018. To be included, these beneficiaries must not have had a claim for an ED visit, observation stay, or inpatient admission with the same diagnosis during the prior 12 months. Among these beneficiaries, 702 (76 percent) had a diagnosis of heart attack and 218 (24 percent) had a diagnosis of stroke. The evaluation drew the 2,247 matched comparison beneficiaries from 33 CAHs in neighboring Nebraska during the same period. Of the comparison beneficiaries, 1,822 (81 percent) had a diagnosis of heart attack and 425 (19 percent) had diagnosis of stroke. Appendix B, Table B.1 displays balance of baseline characteristics between the treatment and matched comparison groups.

Characteristics of treatment and comparison group beneficiaries

A comparison of treatment and comparison group beneficiaries showed that the two groups were well balanced across most of the relevant baseline characteristics (Table 3). The average age of treatment and comparison group beneficiaries during the baseline year was 76 years, and only 2 percent of them were non-White. About 13 percent of both groups were dually eligible for Medicare and Medicaid. The mean hierarchical condition category (HCC) scores for both treatment and comparison beneficiaries were 50 percent higher than the national average for Medicare beneficiaries. Although Medicare expenditures and service use during the year before enrollment were somewhat higher for the treatment group, only the difference in the average number of hospital admissions at baseline differed statistically from zero at the 0.10 level. Appendix B presents the full balance results.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

Measure	Treatment (N = 920)	Comparison (N =2,247)
Demographics		
Age at enrollment, years	76	76
Male, %	47	47
White, %	98	98
Medicare/Medicaid dual status, %	13	12
Health status		
AMI, %	76	76
HCC score[a]	1.5	1.5
Service use and expenditures during the year before enrollment		
Number of hospital admissions (per 1,000 beneficiaries)	398	335
Number of ED visits (per 1,000 beneficiaries)	823	753
Total Medicare expenditures ($ PBPM)	1,123	1,012

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2015, to February 28, 2018, as of August 12, 2019.

Notes: The evaluation defined the baseline year as the 365 days before each episode. It defined the episode initiation as the date of a participant's appearance at a CAH with stroke or AMI diagnosis. Unless otherwise noted, it measured all beneficiary characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

Among all baseline characteristics between treatment and comparison groups, only the difference in number of hospital admissions differed statistically from zero at the 0.10 level, two-tailed test.

Appendix B presents the full balance results. Exact matching variables included diagnoses (AMI or stroke) and quarter of enrollment.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

AMI = acute myocardial infarction; CAH = critical access hospital; ED = emergency department; HCC = hierarchical condition category; PBPM = per beneficiary per month.

Analytic approach

The impact estimates relied on a post-implementation, cross-sectional analysis of the treatment population and a matched group of comparison beneficiaries from neighboring Nebraska. The estimates measured impacts as the difference in outcomes between treatment and comparison beneficiaries during the post-implementation period only. Models were run separately for beneficiaries who presented with stroke and heart attack and controlled for observed characteristics, including beneficiary demographics, dual Medicare–Medicaid eligibility, HCC score, baseline service use and expenditures, and quarter of enrollment. This approach requires that differences on observable variables will capture any existing differences on unobservable characteristics related to outcomes. It assumes that outcomes would not differ between treatment and comparison beneficiaries in the absence of the program. The outcomes for this study were total and inpatient Medicare expenditures, number of hospital admissions, number of ED visits,

and 30-day mortality rates. These outcomes relate directly to the program goals of improving survival and reducing unnecessary service use and costs associated with heart attack and stroke. The study measured outcomes over 6- and 12-month periods after the beneficiary first presented at a CAH with an eligible condition. Appendix A describes the statistical models used to estimate the effects of the program.

IMPACT RESULTS

The analysis suggests the acute care arm of the program led to an estimated **27 percent** lower inpatient expenditures for treatment beneficiaries with a stroke diagnosis during the first year after enrollment (Table 4). However, as described in the limitations, below, inpatient expenditure differences can stem from differences in cost-based reimbursements provided to CAHs. Estimated total Medicare expenditures among treatment beneficiaries with stroke were 9 percent lower than for comparison beneficiaries during the first year after enrollment, the result was not statistically significant. Estimated total and inpatient expenditures among treatment beneficiaries with a heart attack were lower than among their comparison group counterparts as well, but the results were not statistically significant. The program did not have a discernible impact on the number of hospital admissions or ED visits among stroke or heart attack patients. Nor did the program have a statistically significant impact on 30-day mortality for either group of treatment beneficiaries. Appendix C presents the full results of the impact analysis. Appendix D shows the results from the Bayesian analysis.

Table 4. Estimated impact of the Collaborative's acute care arm on selected outcomes

	Heart attack		Stroke	
	1–6 months	1–12 months	1–6 months	1–12 months
Total expenditures ($ PBPM)				
Impact	-$193	-$109	-$787	-$345
Percentage impact	-5%	-4%	-13%	-9%
p-value	0.43	0.52	0.22	0.39
Inpatient expenditures ($ PBPM)				
Impact	-$174	-$115	-$695**	-$374**
Percentage impact	-11%	-11%	-30%	-27%
p-value	0.25	0.23	0.03	0.05
Number of hospitalizations, per 1,000 beneficiaries				
Impact (rate)	4.7	32	4	84
Percentage impact	< 1%	3%	< 1%	6%
p-value	0.96	0.62	0.99	0.53
Number of ED or observation visits, per 1,000 beneficiaries				
Impact (rate)	-11	-40	-70	-37
Percentage impact	< 1%	-2%	-2%	-2%
p-value	0.93	0.65	0.77	0.81

Table 4 (*continued*)

	Heart attack		Stroke	
	1–6 months	1–12 months	1–6 months	1–12 months
30-day mortality				
Impact (rate)	0.29		-3.8	
Percentage impact	6%		-26%	
p-value	0.79		0.21	
Sample size				
Treatment	702	702	218	702
Comparison	1,822	1,822	425	1,822

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2015, to February 28, 2018, as of August 12, 2019.

Note: Impact estimates relied on the regression-adjusted difference between the treatment and inverse propensity weighted comparison group observations. Percentage impacts were then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post-intervention period minus the impact estimate). Appendix C presents full impact estimates. Appendix D shows the results from the Bayesian analysis.

**Significantly different from zero at the .05 level, two-tailed test.

ED = emergency department; PBPM = per beneficiary per month.

Awardee leaders hypothesized that evidence-based protocols, provider education, telemedicine, and TCM and CCM services would collectively improve rural Kansans' heart health and post-stroke survival and reduce total cost of care related to heart disease and stroke. Examining outcomes among all Medicare FFS beneficiaries eligible for the acute care arm of the program, the study found no statistical evidence of reductions in total Medicare spending, rates of hospital admissions, rates of ED visits, or mortality. Low provider engagement and the unwillingness of clinicians to adopt the enhanced acute care arm protocols early in the program might explain the lack of statistically significant findings over the study period. Although the Collaborative ultimately implemented the program successfully, clinicians at the participating CAHs needed training in the acute care protocols and some physicians resisted early in the program. If only a small proportion of these patients received additional ambulatory arm services, which aimed to help patients better manage their conditions, this could also explain the lack of significant findings. Finally, the small number of beneficiaries in the study sample, particularly for the heart attack cohort, made it difficult to detect impacts of reasonable magnitude.

CONCLUSION

The results of the study suggest that the acute care arm of the program had a favorable impact on inpatient Medicare expenditures for beneficiaries with a stroke diagnosis, though not for those with a diagnosis for heart attack. Differences in average hospital payments might have driven an estimated reduction in inpatient expenditures without a discernible reduction in the number of admissions. Because CAHs implemented the acute care arm, for which Medicare bases reimbursement amounts on hospital costs specific to each CAH, the difference in inpatient expenditures for the same type of admission likely stemmed from differences in average hospital costs rather than changes in inpatient admissions. In sensitivity analyses (not shown), the impact

evaluation did not find a discernible impact on average length of stay, further suggesting that differences in average costs between treatment and comparison CAHs, rather than differences in inpatient care, drove the main finding of this study.

Limitations of evaluation

The impact analysis has several limitations. First, the sample relied on all beneficiaries eligible for the acute care arm as identified in Medicare claims data. The impact study was not able to identify beneficiaries treated using the acute care arm protocols. Nor could it distinguish between those who received services under the acute care arm only and those who also received services under the ambulatory care arms. Because the study could not identify those who received the acute care arm protocols, the evaluation averaged the effects over all treatment-eligible FFS beneficiaries who received varying levels of treatment (including no treatment), which leads to an underestimation of the true effect. Second, the evaluation had limited power to detect effects of reasonable sizes. Given the study sample size, minimum detectable effects of program impacts were commonly 15 percent or more of the comparison population mean, requiring changes equal to or greater than 15 percent to statistically identify program effects. Finally, because Medicare FFS payments to CAHs relied on costs, payments for the same types of patients will vary between hospitals in the study. As a result, either treatment effects or differences in costs between treatment and comparison CAHs could have influenced inpatient expenditures, an outcome with statistically significant findings for patients with stroke. The study does not adjust for costs or other hospital-level characteristics.

PROGRAM SUSTAINABILITY

Before its award ended in August 2018, the Collaborative continued its program with four modifications: (1) changing the program name from the Kansas Stroke and Heart Collaborative to the Care Collaborative; (2) expanding the eligibility criteria to include a wider range of chronic conditions (sepsis and trauma); (3) centralizing training of health coaches to improve the efficiency of training and peer learning; and (4) improving the efficiency of delivering health coaching services by risk-stratifying patients to identify those who would most benefit from home visits (as opposed to telephone visits only).

The University of Kansas Health System sustained and expanded the Collaborative program to additional sites by making program participation a requirement of joining the Medicare Shared Savings Plan (MSSP) Accountable Care Organization (ACO). As a result, all participating sites had signed agreements to continue the program after the award period ended. The ACO will

support their participation in the program by providing access to health coaches and support with the program's performance improvement activities.

The awardee calculated that the program cost $83 per member per month, of which its payment model of billing Medicare TCM and CCM codes covered half. The awardee generated the remainder of the necessary funds through shared savings from the ACO, as well as a grant from the United Methodist Hospital Ministries and internal funding from the University of Kansas Health System. However, the awardee sought additional efficiencies and diverse sources of funding to sustain the program in the long term and so, after its award ended, it continued discussions with three Medicaid managed care organizations and other payers about paying for CCM services.

Collaborative's proposed payment model

The University of Kansas Health System implemented two payment approaches to fund Collaborative services after the end of the cooperative agreement: (1) FFS billing Medicare for care management using the TCM and CCM codes and (2) using shared savings from its MSSP ACO. The awardee created an MSSP ACO during the award period and required all hospitals that joined the ACO to participate in the Care Collaborative program.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The analytic approach compared the awardee's eligible beneficiaries to a propensity score matched control group. It then estimated program effects using a regression model of the following form:

(1) $$Y_i = \alpha + \theta Treatment_i + \beta' X_i + \varepsilon_i$$

where Y_i is the outcome of individual *i* in period *t* (for example, total monthly Medicare expenditures during the *t*-th time period since he or she enrolled); α is a constant term; $Treatment_i$ is an indicator for whether the individual was assigned to the group that received program services; X_i are beneficiary characteristics including age, gender, race, dual Medicare–Medicaid eligibility, hierarchical condition category score, baseline values of the outcome variable, total expenditures, and rates of hospitalization and outpatient emergency department visits including observational stays, service use and expenditures, and quarter of enrollment. ε_i is a random disturbance term.

The analysis estimated Equation (1) for each period of interest—6, 12, or 24 months—using separate regressions. The key parameter of interest is θ, which measures the impact of the program for eligible beneficiaries. Appendix C provides results over all outcomes included in the analyses. Appendix A of Volume I of this report provides details on the modeling strategy and the standard set of core outcomes used for this evaluation.

This page has been left blank for double-sided copying.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Table B.1 shows the variables used for matching. It displays the weighted means of baseline characteristics for the 920 treatment beneficiaries and the 2,247 matched comparison beneficiaries used in the impact analysis. The table shows the means, difference in means, the percentage difference, and the standardized difference for each variable. The evaluation calculated the standardized difference as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit. The matching variables included Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category [HCC] score); number of hospital admissions (12 months and 30 days before episode); number of emergency department (ED) or observation visits (12 months and 30 days before episode); number of primary care visits (12 months and 30 days before episode); ED or observation visit on the day of enrollment; hospital admission on the day of enrollment; and Medicare expenditures in total (12 months and 30 days before episode). The evaluation used inverse propensity score weights for the comparison episodes. It required an exact match for acute myocardial infarction (AMI) diagnosis at the enrollment encounter and the quarter of enrollment for treatment beneficiaries and at the pseudo-enrollment date for the comparison beneficiaries. It then measured the variables over various specified intervals within the 12 months before enrollment in the intervention.

The table also shows the results of the equivalency-of-means tests. The *p*-values come from a weighted two-sample t-test, which provides evidence of a statistically significant difference in the means. The equivalence test *p*-values are the greater of the two one-sided weighted t-test *p*-values equivalence test, which assesses whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. Finally, the evaluation performed an omnibus test in which the null hypothesis was that the treatment and matched comparison groups were balanced across all linear combinations of the covariates. It used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely associated with study outcomes. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

Table B.1. Baseline characteristics of the Collaborative's treatment and matched comparison groups

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test p-value	Equivalence p-value
Demographics							
Age, years	76 (0.37)	76 (0.23)	0.02 (0.51)	< +/-1	0.00	0.97	< 0.01
Male, %	47 (1.6)	47 (1.1)	0.00 (2.3)	< +/-1	0.00	1.00	< 0.01
Female, %	53 (1.6)	53 (1.1)	0.00 (2.3)	< +/-1	0.00	1.00	< 0.01
White, %	98 (0.43)	98 (0.24)	-0.06 (0.62)	< +/-1	0.00	0.92	< 0.01
Black, %	0.11 (0.11)	0.06 (0.06)	0.05 (0.14)	42	0.01	0.75	< 0.01
Hispanic, %	0.54 (0.24)	0.46 (0.15)	0.08 (0.33)	15	0.01	0.80	< 0.01
American Indian, Alaska Native, Asian or Pacific Island American, or other, %	0.54 (0.24)	0.60 (0.11)	-0.05 (0.35)	-10	-0.01	0.88	< 0.01
Unknown, %	0.54 (0.24)	0.55 (0.13)	-0.01 (0.36)	-2.0	0.00	0.98	< 0.01
Medicare entitlement and dual eligibility status, %							
Dually eligible for Medicare and Medicaid	13 (1.1)	12 (0.70)	0.83 (1.6)	6.3	0.02	0.60	< 0.01
Health status and diagnoses							
HCC score[a]	1.5 (0.04)	1.5 (0.02)	0.05 (0.06)	3.2	0.04	0.39	< 0.01
Medicare expenditures							
Total expenditures	1,123 (61)	1,012 (31)	111 (82)	9.9	0.06	0.17	< 0.01
Total expenditures, 3 months before enrollment	1,313 (103)	1,217 (57)	96 (148)	7.3	0.03	0.52	< 0.01
Service utilization							
Hospitalizations	398 (27)	335 (14)	63 (37)	16	0.08	0.09	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test p-value	Equivalence p-value
Total hospitalizations, 3 months before enrollment	470 (46)	411 (28)	59 (68)	12	0.04	0.39	< 0.01
Hospital admission, day of enrollment	28 (1.5)	28 (0.91)	0.04 (2.1)	< +/-1	0.00	0.98	< 0.01
ED or observation visit, day of enrollment	75 (1.4)	75 (0.82)	0.26 (2.0)	< +/-1	0.01	0.90	< 0.01
Primary care visits, any setting	9,182 (271)	8,794 (154)	388 (384)	4.2	0.05	0.31	< 0.01
Primary care visits, any setting, 3 months before enrollment	10,222 (385)	9,930 (239)	292 (595)	2.9	0.02	0.62	< 0.01
Total outpatient ED or observation visits	823 (45)	753 (28)	69 (64)	8.4	0.05	0.28	< 0.01
Outpatient ED or observation visits, 3 months before enrollment	1,104 (85)	993 (51)	111 (121)	10	0.04	0.36	< 0.01
Propensity score	0.27 (0.00)	0.27 (0.00)	0.00 (0.00)	< +/-1	0.02	0.65	< 0.01
Number of beneficiaries	920	2,247					
Omnibus test				Chi-squared statistic 25.83	Degrees of freedom 19.00	*p*-value 0.13	

Source: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2015, to February 28, 2018, as of August 12, 2019.

Notes: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include diagnoses (AMI or stroke) and quarter of enrollment.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

AMI = acute myocardial infarction, CAH = critical access hospital; ED = emergency department; HCC = hierarchical condition category; SE = standard error.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates

This page has been left blank for double-sided copying.

Table C.1 shows the impact estimates for the stroke and acute myocardial infarction interventions for the full study population. The evaluation estimated the models over Medicare expenditures and number of services used (per 1,000 beneficiaries) in total and by type of service, as well as for the probabilities of hospital readmission and mortality within 30 and 90 days of enrollment. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impacts of the Collaborative's stroke and AMI interventions on select Medicare FFS expenditures (dollars PBPM) and use measures during 6-, 12-, and 24-month follow-up periods for CAHs

	UKS stroke intervention					UKS AMI intervention				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total Medicare expenditures ($ PBPM)[b]										
Baseline year	951	936				1,152	991			
Months 1 to 6	5,356	6,143	-787 (642)	-13%	0.22	3,479	3,672	-193 (247)	-5.2%	0.43
Months 1 to 12	3,528	3,874	-345 (401)	-8.9%	0.39	2,622	2,730	-109 (168)	-4.0%	0.52
Months 1 to 24	2,669	2,949	-279 (291)	-9.5%	0.34	2,158	2,272	-114 (126)	-5.0%	0.36
Acute inpatient expenditures ($ PBPM)[b]										
Baseline year	239	276				291	293			
Months 1 to 6	1,596	2,290	-695** (320)	-30%	0.03	1,420	1,595	-174 (153)	-11%	0.25
Months 1 to 12	1,016	1,389	-374** (188)	-27%	0.05	978	1,093	-115 (96)	-11%	0.23
Months 1 to 24	762	1,027	-264** (130)	-26%	0.04	756	867	-112 (69)	-13%	0.10
Hospital stays, per 1,000 beneficiaries										
Baseline year	339	277				405	340			
Months 1 to 6	2,213	2,209	4.0 (215)	< 1%	0.99	1,413	1,409	4.7 (92)	< 1%	0.96
Months 1 to 12	1,437	1,354	84 (135)	6.2%	0.53	1,023	991	32 (63)	3.2%	0.62
Months 1 to 24	1,090	1,003	87 (100)	8.6%	0.39	816	795	21 (49)	2.6%	0.67
ED or observation visits, per 1,000 beneficiaries										
Baseline year	601	594				870	767			
Months 1 to 6	2,883	2,953	-70 (235)	-2.4%	0.77	3,100	3,111	-11 (116)	< 1%	0.93
Months 1 to 12	1,924	1,961	-37 (155)	-1.9%	0.81	2,174	2,215	-40 (90)	-1.8%	0.65
Months 1 to 24	1,531	1,486	45 (123)	3.0%	0.72	1,731	1,773	-43 (75)	-2.4%	0.57
Beneficiary with readmission										
Baseline year	2.8	1.8				3.6	2.6			
Months 1 to 6	13	12	1.2 (3.1)	9.9%	0.70	8.9	8.2	0.78 (1.3)	9.6%	0.56
Months 1 to 12	15	13	2.1 (3.3)	16%	0.53	12	11	1.4 (1.5)	13%	0.35
Months 1 to 24	20	17	2.5 (3.8)	15%	0.51	17	17	0.90 (1.9)	5.5%	0.63

Table C.1 (*continued*)

	UKS stroke intervention					UKS AMI intervention				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
30-day mortality										
30 days	11	14	-3.8 (3.0)	-26%	0.21	5.6	5.3	0.29 (1.1)	5.5%	0.79
90-day mortality										
90 days	16	20	-4.0 (3.4)	-20%	0.24	7.5	7.8	-0.24 (1.2)	-3.1%	0.84
Sample sizes										
Number of beneficiaries										
Baseline year	218	425				702	1,822			
Months 1 to 6	218	425				702	1,822			
Months 7 to 12	218	425				631	1,673			
Months 13 to 18						543	1,437			
Months 19 to 24						393	1,057			
Months 1 to 12						702	1,822			
Months 13 to 24						544	1,438			
Months 1 to 24	218	425				702	1,822			

Source: Mathematica's analysis of information from Medicare claims and enrollment data as of August 12, 2019.

Note: Impact estimates for expenditures and number of visits or stays relied on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any 30-day and 90-day mortality, as well as any beneficiary with readmission, is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

[b] 98th percentile values for top-coding were determined from the weighted distribution of treatment beneficiaries pooled over the four semiannual periods covering the baseline and follow-up years.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

AMI = acute myocardial infarction; CAH = critical access hospital; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month, SE = standard error; UKS = University of Kansas Health System.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the analysis also estimated program impacts for the University of Kansas Health System (UKS) using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. Drawing probabilistic conclusions requires external or prior evidence. In this analysis, the findings from the evaluation of 87 awardees included in the Round 1 of the Health Care Innovation Awards provided the prior evidence, with more weight on results from awardees with background characteristics similar to those of UKS. The analysis calculated probabilities using the results of a Bayesian regression that jointly models impacts on three core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for three core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. As in the body of the report, impact estimates focus on beneficiaries receiving care at critical access hospitals and separately assess patients with acute myocardial infarction (AMI) and stroke. Combining prior evidence from Round 1 of the Health Care Innovation Awards with the estimates from the frequentist regression for UKS led to a Bayesian estimate of the program's impact on total Medicare expenditures of 2 percent (an estimated increase of $63 per beneficiary per month) for AMI patients and an impact of 4 percent (an estimated increase of $169 per beneficiary per month) for stroke patients.

Table D.1. Comparison of frequentist and Bayesian impact estimates for UKS in the first year after enrollment

		Impact estimate (95 percent interval)		Percentage impacts		
Subgroup	Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
AMI	Total expenditures ($ PBPM)	-109 (-438, 221)	63 (-150, 275)	6%	-4%	2%
	Hospital admissions	32 (-92, 156)	26 (-53, 103)	6%	3%	3%
	ED visits	-40 (-218, 137)	46 (-128, 222)	6%	-2%	2%
Stroke	Total expenditures ($ PBPM)	-345 (-1,131, 441)	169 (-216, 537)	7%	-9%	4%
	Hospital admissions	84 (-180, 347)	62 (-75, 191)	8%	6%	5%
	ED visits	-37 (-340, 266)	80 (-119, 268)	7%	-2%	4%

Sources: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2015, to February 28, 2018, as of August 12, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

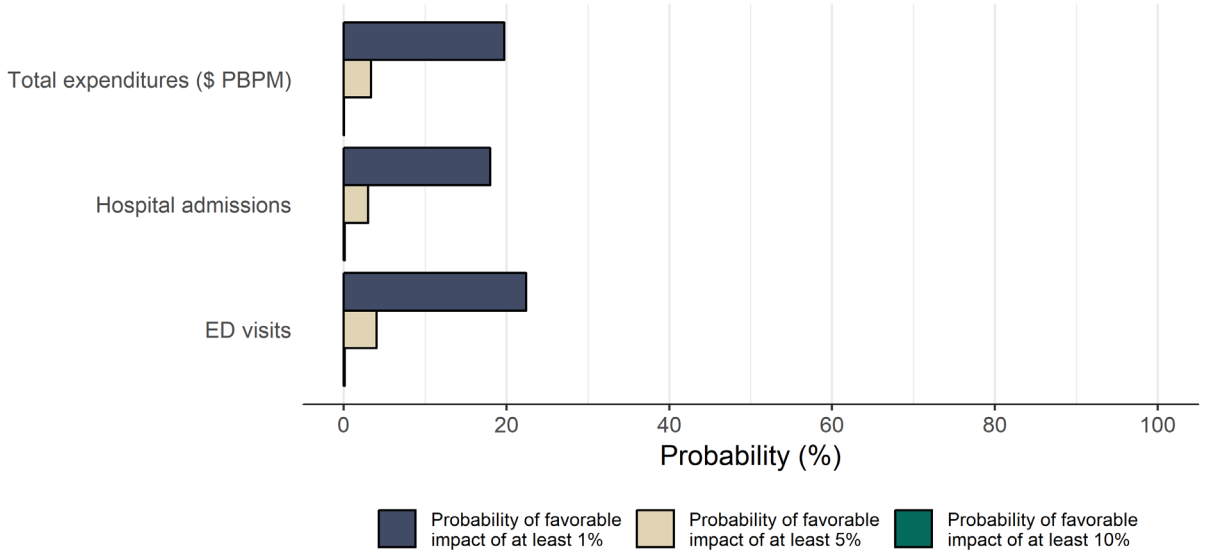
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

AMI = acute myocardial infarction; ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month; UKS = University of Kansas Health System.

Because the frequentist results relied on a small sample and are therefore imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian and frequentist results substantively agree in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figures D.1 and D.2 show the probability that UKS achieved favorable impacts for each subgroup of interest during the first year on three core outcomes at three different thresholds: a favorable impact of (1) 1 percent or more, (2) 5 percent or more, and (3) 10 percent or more.

Figure D.1. Probability that the UKS program had a favorable impact on key outcomes, AMI subgroup

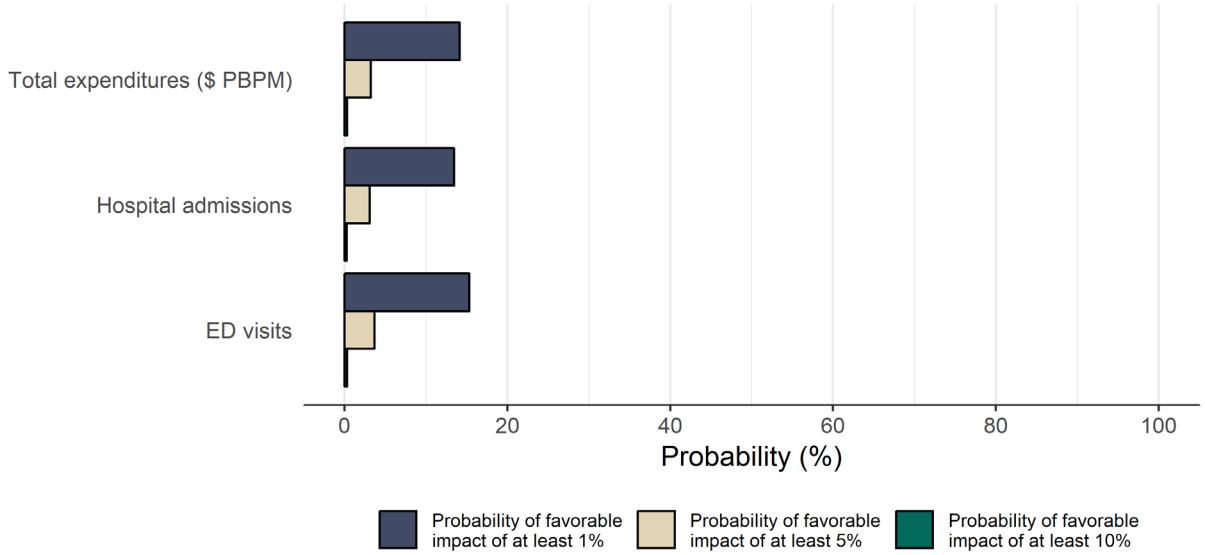


Sources: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2015, to February 28, 2018, as of August 12, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions relied on data from the HCIA R1 evaluation.

AMI = acute myocardial infarction; ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month; UKS = University of Kansas Health System.

Figure D.2. Probability that the UKS program had a favorable impact on key outcomes, stroke subgroup



Sources: Mathematica's analysis of Medicare claims and enrollment data from March 1, 2015, to February 28, 2018, as of August 12, 2019. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

ED = emergency department; HCIA R1 = Round 1 of the Health Care Innovation Awards; PBPM = per beneficiary per month; UKS = University of Kansas Health System.

There is a small probability—in the range of 20 percent for AMI patients and closer to 15 percent for stroke patients—that the UKS program had a favorable impact of 1 percent or more on total Medicare expenditures, hospital admissions, and emergency department visits. These probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the UKS program did not have a meaningful impact on total expenditures or service use.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: University of New Mexico

September 2020

Suhui (Evelyn) Li, Ellie Coombs, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

UNIVERSITY OF NEW MEXICO

The University of New Mexico (UNM) used funding from the Round 2 of the Health Care Innovation Awards to launch the Access to Critical Cerebral Emergency Support (ACCESS) program. Because of a lack of access to neurosurgeons and neurologists in rural New Mexico emergency departments (EDs), local hospitals unnecessarily transferred many patients with neuro-emergent conditions, such as mild traumatic brain injuries or strokes, to tertiary care providers—such as UNM—for diagnosis and treatment. The goal of the ACCESS program was to reduce unnecessary and costly transfers of patients by supporting rural hospitals in effectively diagnosing and treating neuro-emergent conditions through telehealth consultations with neurologists and neurosurgeons.

Important issues for understanding the evaluation

- The ACCESS program aimed to help rural hospitals efficiently treat ED patients with neuro-emergent conditions and thus reduce unnecessary and costly transfers to tertiary care hospitals.
- ED physicians enrolled patients into the ACCESS program; they made telehealth referral decisions based on the patient's *presenting* symptoms upon arrival to the ED and medical history.
- Because it was not possible to use Medicare claims data to identify a comparison group that was similar to the intervention group at the time of presentation to the ED, this study does not present impact estimates.

Under the ACCESS program, when a patient presented in a participating hospital's ED with a potential neuro-emergent condition, the ED physician used the Net Medical Xpress Solutions (NMXS) telehealth platform to communicate with a neurologist or neurosurgeon. These specialists, contracted by either NMXS or UNM, examined the patient, reviewed imaging, and discussed treatment options through the technology's secure file transfer and video capabilities. Based on the consultation, the specialist discharged, admitted, or transferred the patient. Telehealth coordinators at each participating hospital facilitated the consultation process and served as the primary liaison between ACCESS and ED staff at participating hospitals. The awardee hypothesized that telehealth consultations would decrease the time it took for the patient to receive a treatment recommended by a specialist, decrease unnecessary hospital transfers, improve physicians' confidence in treatment decisions, and improve patients' satisfaction because they received treatment closer to home. In turn, access to specialty care provided locally through telehealth would result in better health outcomes and lower health care costs. Table 1 summarizes the program's key characteristics.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The awardee implemented the ACCESS program to facilitate telehealth consultations for patients who presented at a participating hospital's ED with a neuro-emergent condition.
Major innovation	Telehealth is an innovative approach for providing specialty care to people in rural areas. Lack of access to specialists means that patients must often travel long distances to needed care. ACCESS telehealth fills this gap by making specialists available to patients and their providers through innovative technology, including video monitors, diagnostic equipment, and scan-sharing capabilities. ACCESS also provides innovation through telehealth coordinators who act as program advocates within hospitals and UNM clinical staff who build a hospital's capacity to treat neuro-emergent conditions through training.
Program components	The program used telehealth consultations between the treating physician in spoke hospitals with consulting neurologists and neurosurgeons in the hub hospital.
Target population	The program sough to engage adults and children who presented to a participating ED with a neuro-emergent condition. Adults might be insured through Medicare, Medicaid, or not covered by either insurance. Children might be insured through Medicaid, the Children's Health Insurance Program, or neither.
Total enrollment	The awardee reported that from May 2015 to May 2018, the ACCESS program provided consultations to 2,545 patients enrolled in Medicare or Medicaid, which represents 30 percent of its original three-year enrollment target of 8,504 consultations.
Theory of change or theory of action	The awardee hypothesized that telehealth consultations would decrease the time it took for the patient to receive a treatment recommended by a specialist, decrease unnecessary hospital transfers, improve physicians' confidence in treatment decisions, and improve patients' satisfaction because they received treatment closer to home. In turn, the access to specialty care provided locally through telehealth would result in better health outcomes and lower health care costs.
Award amount	$15,042,466
Effective launch date	May 4, 2015
Program settings	Hospital ED
Market area	Initially rural, then expanded to the entire state
Market location	New Mexico
Target outcomes	• Lower health care costs • Reduced unnecessary hospital transfers
Payment model	Billing existing Medicare FFS codes for neurological and neurosurgery telehealth services
Sustainability plans	After its award ended in August 2018, 18 hospitals had signed contracts to continue participating in UNM's ACCESS program, including 15 hospitals in which the program was fully operational and 3 hospitals that were still implementing the program. By the end of the award, the awardee had not executed contracts with commercial payers that would cover telehealth consultations for non-Medicare FFS patients, although the negotiations were still underway. The awardee hoped to expand the program to Medicaid patients, given the state's Medicaid program reported plans to issue a directive to its managed care organizations to cover consultations within their current capitated payments.

ACCESS = Access to Critical Cerebral Emergency Support program; ED = emergency department; FFS = fee-for-service; UNM = University of New Mexico.

A rigorous impact evaluation was not possible for this awardee because of how it identified and enrolled eligible participants. Physicians enrolled patients into the ACCESS program; they made telehealth referral decisions based on the patient's *presenting* symptoms upon arrival to the ED and medical history. However, using Medicare claims, less than half of ED claims from ACCESS-participating hospitals included a presenting diagnosis determined upon the patient's arrival at the ED (the "reason for visit" diagnosis code). In addition, principal diagnosis at discharge (determined after medical evaluation and treatment) could not serve as a proxy for presenting diagnosis because the telehealth consultation itself likely influenced the diagnosis. Evidence suggests only weak correlation between the presenting and principal diagnoses at discharge among program participants. Because of the lack of data on presenting diagnosis and the risk of selection bias associated with using principal diagnosis at discharge, the analysis could not use Medicare claims data to identify comparison beneficiaries similar to the program participants. As a result, this report describes only the demographic and health characteristics of Medicaid participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis. Appendix Table A.1 describes the identification of the study sample.

Table 2. Key features of descriptive analysis

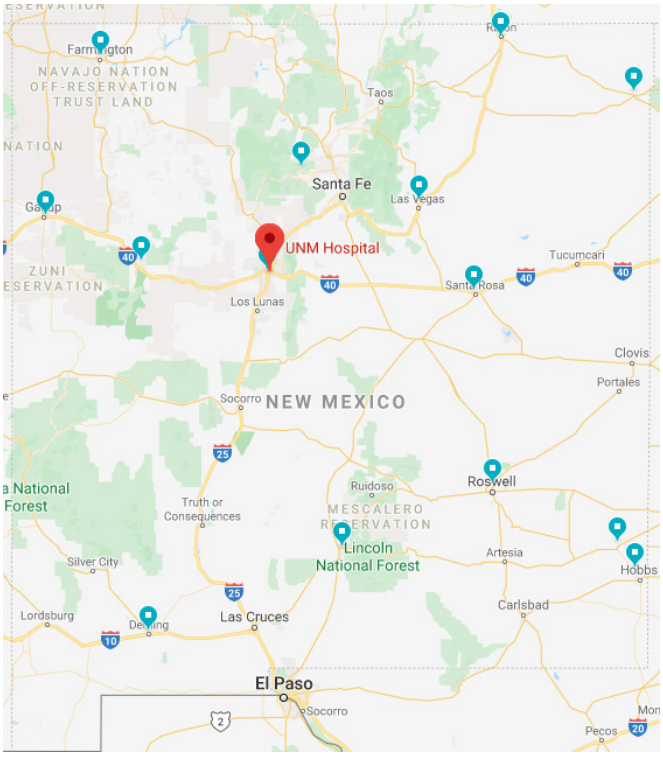
Features	Description
Descriptive analysis	Because of how the program identified and enrolled participants, it was not possible to conduct a rigorous impact evaluation of this program. Health care claims could not accurately replicate the identification of comparison beneficiaries similar to the intervention participants because ED physicians enrolled participants based on presenting illness upon arrival at the ED rather than final diagnosis. As a result, this report describes only the demographic and health characteristics of Medicare FFS participants.
Intervention group for descriptive analysis	Of the 2,545 program participants who enrolled in the ACCESS program between May 4, 2015 and May 31, 2018, 1,256 linked to the Medicare enrollment database. Beneficiaries were not included in the descriptive analysis if they did not have an ED claim from one of the ACCESS hospitals during the program period (71 beneficiaries) or did not meet met standard claims-based inclusion criteria (73). The descriptive analysis presented in this report was limited to 1,112 Medicare FFS beneficiaries.
Limitations	Due to the problems noted above, the analysis cannot be used to make inferences about the impact of this program on Medicare costs or other program outcomes.

ACCESS = Access to Critical Cerebral Emergency Support; FFS = fee-for-service;

PROGRAM DESIGN AND ADAPTATION

The awardee conducted extensive outreach to encourage hospitals to participate in the ACCESS program, contacting all hospitals in New Mexico. Interested hospitals underwent a lengthy process to join the program, which involved introductory communications and conference calls, site visits from the awardee, and contract negotiations. Participating hospitals then credentialed the neurologists and neurosurgeons who provided ACCESS consultations to the ED physicians at the spoke hospitals. To implement the program, the awardee installed the NMXS technology and trained the hospital's telehealth coordinator and ED staff on making a request and facilitating the

Figure 1. Locations of hospitals participating in ACCESS program



consultation. Over the program period, the awardee streamlined the consultation request process in three ways. First, it simplified the online request form. Second, the awardee helped hospitals categorize acuity to prioritize urgent cases. Third, it trained hospitals on the difference between neurology and neurosurgery to enable them to make the appropriate requests.

The awardee received a 12-month no-cost extension through August 30, 2018. By the end of the no-cost extension period, 15 hospitals had started seeing patients through the ACCESS program, and 3 more had signed contracts. As Figure 1 shows, most of the 15 participating spoke hospitals were in rural New Mexico, at least 100 miles from the hub provider, UNM. Some participating hospitals had already used the NMXS telehealth technology before the ACCESS program. Under the cooperative agreement, the awardee helped these hospitals upgrade the technology, added the neurosurgery component, and provided training and support in providing neurological care.

After the program launched, when a patient presented in a participating hospital's ED with a possible neuro-emergent condition, the ED clinician would request a telehealth consult through the NMXS online platform, along with relevant computed tomography (CT) scans. The specialist would then connect with the ED to conduct a retinal scan, consult with the ED clinician, and interview the patient. The specialist then recommended a treatment plan, which included advice on discharging, admitting, or transferring the patient.

Patients could decline the telehealth service offered by the ED staff, but most of them and their families chose to consult with the specialist through the telehealth technology. According to the awardee, although the ACCESS program focused exclusively on patients with neurological emergencies, not all enrolled patients had a discharge diagnosis related to neuro-emergent conditions. For some patients with altered mental status who received a telehealth consultation through the ACCESS program, subsequent examinations found non-neurological conditions caused their presenting symptoms, and hospitals reported the latter diagnosis at discharge.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee faced challenges meeting its enrollment goal. From May 2015 (when it launched its program) through August 2017, the ACCESS program provided telehealth consultations to 1,450 patients enrolled in Medicare or Medicaid. The number of consultations grew to 2,545 by May 2018, which still represented only 30 percent of its original three-year projection of 8,540 participants. Consultations varied from an average of 20 per month at larger hospitals to 1 or 2 per month at smaller hospitals.

A main factor leading to fewer consultations than anticipated was the difficulty with recruiting hospitals. Despite the awardee's concerted efforts to recruit hospitals, many hesitated to join ACCESS because of concerns that they could not afford the program when the cooperative agreement ended, or they had plans to adopt alternative, less expensive telehealth planforms. For example, instead of paying a fee for every consultation with ACCESS ($600 to $1,200 depending on the type and length of consultation), larger rural hospitals aimed to pay an annual subscription fee, which might be more cost-effective for higher-use settings. Furthermore, staff turnover within rural New Mexico hospitals led to the awardee making multiple recruitment attempts per hospital, further delaying recruitment. The awardee noted that "there is so much turnover in these hospitals that by the time we got something worked out with the CEO [chief executive officer], he was gone and we had to start all over."

Even after hospitals signed contracts, the awardee experienced several barriers to implementing the ACCESS program among participating hospitals. First, credentialing the neurologists and neurosurgeons providing consulting services to the participating hospitals could take months, delaying the launch of the ACCESS program. Second, the awardee struggled to identify new ED clinical staff who needed ACCESS training and to ensure they were aware of the NMSX telehealth technology. Staff turnover might have been the reason only half the hospital staff surveyed by the evaluation team had attended a formal ACCESS training. Although bringing new staff up to speed on the program was the role of the telehealth coordinators, some had multiple roles in these small rural hospitals and could not dedicate the time required to the project. This meant that when UNM staff saw use trailing off, they had to follow up with hospitals to ensure that new staff were aware of the technology and started using it. Third, the small sizes of some participating hospitals slowed their adoption of the new technology. In some participating EDs with only one or two neuro-emergent cases a month, the time lag between consultations could mean that ED staff forgot about or deprioritized the process.

Despite challenges with recruiting hospitals that led to fewer consultations than anticipated, hospital staff reported that the consultation process itself was straightforward and enabled them to connect their patients to needed specialty care. UNM and hospital staff indicated through surveys and interviews that the program was a cost-effective method to improve care delivery. Of the 21 nonclinician hospital staff who completed the survey, most reported ACCESS had a

positive effect on the quality (76 percent) and efficiency (81 percent) of care provided. More than 70 percent reported that the program was worth the effort and that they would recommend it to a colleague. Although few clinicians completed the clinician survey, those who did agreed that ACCESS had a positive effect on the quality of care they provided.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Enrolling participants

Of the 2,545 patients the program enrolled, 1,112 were Medicare FFS beneficiaries. Table 3 shows the range of clinical conditions, identified through Agency for Healthcare Research and Quality's Clinical Classifications Software (CCS) categories, based on principal diagnoses at discharge among these Medicare beneficiaries who were enrolled in the ACCESS program through May 2018. Consistent with the stated target of the ACCESS program, the most prevalent principal diagnoses at discharge among participants reflected conditions that likely needed a neurological or neurosurgical consultation. These conditions included acute cerebrovascular disease, transient cerebral ischemia, epilepsy, syncope, other nervous system disorders, and intracranial injury. Most participants (663) had one of these neuro-emergent conditions. A few had a principal diagnosis likely associated with non-emergent neurological conditions, such as headache, malaise, and fatigue. The remaining participants (about 33 percent), however, had diagnoses that spanned a broad range of non-neurological conditions, such as urinary tract infections, septicemia, and diabetes mellitus. Appendix A, Table A.1 describes the identification of the sample used for the descriptive analysis. Appendix B includes a full list of clinical conditions among ACCESS program participants.

Table 3. Conditions based on principal diagnoses at discharge among Medicare FFS participants of the ACCESS program

CCS categories	Number of participants (N = 1,112)	Percentage of participants
Acute cerebrovascular disease	261	23.5
Transient cerebral ischemia	165	14.8
Other nervous system disorders	97	8.7
Epilepsy; convulsions	72	6.5
Syncope	36	3.2
Intracranial injury	32	2.9
Residual codes; unclassified	32	2.9
Malaise and fatigue	31	2.8
Conditions associated with dizziness or vertigo	28	2.5
Headache; including migraine	19	1.7

Table 3 (*continued*)

CCS categories	Number of participants (N = 1,112)	Percentage of participants
Urinary tract infections	18	1.6
Other connective tissue disease	18	1.6
Septicemia	17	1.5
Acute and unspecified renal failure	16	1.4
Spondylosis; intervertebral disc disorders; other back problems	12	1.1
Other fractures	12	1.1
Cardiac dysrhythmias	11	1.0
Other and ill-defined cerebrovascular disease	11	1.0
Diabetes mellitus with complications	10	0.9
Pneumonia (except caused by tuberculosis or STIs)	9	0.8
Other conditions	205	18.4

Source: Mathematica's analysis of Medicare claims and enrollment data from January 2014 to August 31, 2018, as of November 25, 2019.

ACCESS = Access to Critical Cerebral Emergency Support; CCS = Clinical Classifications Software; FFS = fee-for-service; STIs = sexually transmitted infections.

An analysis of Medicare FFS claims from the ED visit in which the consultation occurred revealed that among participants with a presenting diagnosis present on the claim, only 25 percent had a matching principal diagnosis at discharge. Further, treatment beneficiaries were less likely to have the same presenting and principal diagnoses than beneficiaries in the potential comparison group, suggesting the intervention itself can alter the principal diagnosis reported on the discharge claim for the treatment group (and used to identify the comparison group). For example, some patients presenting with malaise and fatigue (and enrolled in the program) were later diagnosed as having a urinary tract infection or cardiac dysrhythmias. Thus, the lack of concordance between presenting and principal diagnoses at discharge would lead to biased impact results.

Characteristics of Medicare FFS participants in the ACCESS program

Among the 1,112 participants enrolled in Medicare FFS, the average age was about 74, with 85 percent older than 65 (Table 4). More than 80 percent of the Medicare participants were White, and 32 percent of the participants were dually eligible for Medicare and Medicaid. Congestive heart failure, vascular disease, and chronic obstructive pulmonary disease were by far the most prevalent chronic conditions among Medicare participants at the baseline. The average hierarchical condition categories risk score of participants was 1.8, indicating that their expected Medicare annual spending was nearly twice that of the average Medicare FFS beneficiary. During the baseline year, these participants had about 471 acute care hospitalizations and 1,362 ED visits per 1,000 beneficiaries per year. Average Medicare spending among participants was $1,445 per beneficiary per month during the baseline year.

Table 4. Baseline characteristics of Medicare FFS beneficiaries

Characteristics	Participants (N = 1,112)
Demographics	
Age at enrollment, years	74
Age group, %	
Younger than 65	15
65 to 74	32
75 to 84	33
85 and older	20
Male, %	44
White, %	84
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	71
Disability insurance benefits[a]	28
Medicare–Medicaid dual status, %	32
Chronic conditions, %	
COPD	21
CHF	24
Morbid obesity	5.8
Vascular disease	22
Major depressive disorder	14
HCC score[b]	
Mean	1.8
25th percentile	0.83
Median	1.4
75th percentile	2.3
Service use and expenditures during the year before enrollment	
Any hospitalizations, %	28
Hospital stays per 1,000 beneficiaries[c]	471
Any ED visits, %	57
ED visits per 1,000 beneficiaries[c]	1,362
Total Medicare expenditures ($ PBPM)	1,445

Source: Mathematica's analysis of Medicare claims from January 2014 through August 31, 2018, as of November 25, 2019.

Notes: The baseline year is the 365 days before and including each beneficiary's enrollment date. The enrollment date is the index ED visit date. All beneficiary characteristics were measured during or as of the end of the baseline year.

The study weighted the statistics for participants to account for the number of months a beneficiary was enrolled in FFS Medicare.

[a] Includes beneficiaries with both a disability and ESRD.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

Table 4 (*continued*)

[c] The number of hospital stays and ED visits was topcoded at the 98th percentile of the participant group's distribution of hospital stays and ED visits during the baseline period.

ACCESS = Access to Critical Cerebral Emergency Support; CMS = Centers for Medicare & Medicaid Services; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; PBPM = per beneficiary per month.

CONCLUSIONS

UNM partly succeeded in implementing the ACCESS program to support effective diagnosis and treatment of ED patients with neuro-emergent conditions at rural hospitals in New Mexico. Implementation challenges included that hospitals were hesitant to participate because of financial concerns about the program, hospitals' staff turnover led to the need for frequent training of new staff, the lengthy credentialing process delayed the start of consultations at participating hospitals, and the small sizes of participating hospitals slowed the adoption of the new technology. The criteria that ED clinicians used to identify and enroll participants into the ACCESS program were not observable for potential comparison cases and were likely to affect patients' outcomes. Thus, it was not possible to select an equivalent comparison group to conduct a rigorous impact evaluation.

PROGRAM SUSTAINABILITY

By the time the award ended in August 2018, 18 hospitals had signed contracts to continue participating in the UNM's ACCESS program. The awardee retained four central staff for the program who continued recruiting hospitals to join the program, supported hospitals' use of the program's technology, and educated hospitals on providing higher quality neuro-emergent care to their patients.

UNM expects that hospitals will sustain the program by billing existing Medicare FFS codes that cover neurological and neurosurgery telehealth services, absorbing the remaining costs through internal funds. The awardee expects that program participation will enable hospitals to retain more patients who otherwise would have been transferred to another hospital; the additional revenues generated by increased retention would help offset the hospitals' costs of operating the program.

By the end of the award, the awardee was unable to execute contracts with commercial payers who would cover telehealth consultations for non-Medicare FFS patients, although the

UNM's proposed payment model

UNM proposed billing existing Medicare FFS codes that cover neurological and neurosurgery telehealth services. The awardee expects that the FFS payments will not fully cover the cost of program consults, and hospitals will have to absorb the remaining costs in the near term. The awardee expects that the program will ultimately pay for itself, however, because implementing the program can help hospitals increase inpatient revenue by retaining patients who previously would have been transferred to another hospital.

negotiations were still underway. The awardee also hoped that it could expand the program to Medicaid patients, given the state's plans to issue a directive to its Medicaid managed care organizations to cover telehealth consultations within their current capitated payments.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total number of Medicare beneficiaries in awardee's finder file as of May 31, 2018		**1,256**
Beneficiaries who did not have an ED claim at any ACCESS hospital during the program period	71	1,185
Beneficiaries who did not meet the standard claims-based inclusion criteria		
Not alive at the time of enrollment	1	1,184
Not enrolled in both Medicare Part A and Part B	33	1,151
Enrolled in Medicare Advantage	3	1,148
Medicare was not primary payer	11	1,137
Lacked 90 days of FFS enrollment during baseline	25	1,112
Final analytic sample		**1,112**

Source: Mathematica's analysis of awardee's finder file and Medicare claims and enrollment data from January 2014 through August 31, 2018, as of November 25, 2019.

ACCESS = Access to Critical Cerebral Emergency Support; FFS = fee-for-service.

This page has been left blank for double-sided copying.

Appendix B

Supplementary tables

This page has been left blank for double-sided copying.

Table B.1. Clinical conditions based on principal diagnoses among ACCESS program Medicare FFS participants

CCS categories	Count (N = 1,112)	Percentage
Acute cerebrovascular disease	261	23.5
Transient cerebral ischemia	165	14.8
Other nervous system disorders	97	8.7
Epilepsy; convulsions	72	6.5
Syncope	36	3.2
Intracranial injury	32	2.9
Residual codes; unclassified	32	2.9
Malaise and fatigue	31	2.8
Conditions associated with dizziness or vertigo	28	2.5
Headache; including migraine	19	1.7
Urinary tract infections	18	1.6
Other connective tissue disease	18	1.6
Septicemia	17	1.5
Acute and unspecified renal failure	16	1.4
Spondylosis; intervertebral disc disorders; other back problems	12	1.1
Other fractures	12	1.1
Cardiac dysrhythmias	11	1.0
Other and ill-defined cerebrovascular disease	11	1.0
Diabetes mellitus with complications	10	1.0
Pneumonia (except that caused by tuberculosis or sexually transmitted infection)	9	0.8
Fluid and electrolyte disorders	9	0.8
Hypertension with complications and secondary hypertension	9	0.8
Late effects of cerebrovascular disease	8	0.7
Delirium dementia and amnestic and other cognitive disorders	8	0.7
Blindness and vision defects	8	0.7
Essential hypertension	8	0.7
Nonspecific chest pain	7	0.6
Paralysis	7	0.6
Other circulatory disease	6	0.5
Superficial injury; contusion	6	0.5
Acute myocardial infarction	5	0.5

Table B.1 (*continued*)

CCS categories	Count (N = 1,112)	Percentage
Open wounds of head; neck; and trunk	5	0.5
Other injuries and conditions due to external causes	5	0.5
Other categories	114	10.3
Total	**1,112**	**100**

Source: Mathematica's analysis of Medicare claims and enrollment data from January 2014 through August 31, 2018, as of November 25, 2019.

ACCESS = Access to Critical Cerebral Emergency Support; CCS = Clinical Classifications Software; CNS = central nervous system; FFS = fee for service.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: The University of North Carolina at Chapel Hill

September 2020

Poonam Pardasaney, Julia Doherty, Bob Schmitz, and Randall Brown

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

The University of North Carolina at Chapel Hill (UNC) received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the Better Back Care (BBC) program in the Raleigh, Durham, and Chapel Hill region. The goal of the program was to implement a new care delivery model for patients with acute (not chronic), nonspecific low back pain (LBP) that emphasized evidence-based, conservative treatments over unnecessary imaging, injections, and surgery. The target population included patients 18 years or older with acute, nonspecific LBP making their first visit to a participating provider. The program began enrolling patients in February 2015 and stopped in August 2017. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- BBC's care delivery model emphasized evidence-based guideline adherence and conservative treatment, when appropriate, for acute, nonspecific low back pain (LBP).
- The program aimed to reduce unnecessary imaging, injections, and surgery and thereby reduce costs of care, improve patient-reported function and general health outcomes, and improve patients' satisfaction.
- This analysis of baseline characteristics relied on 350 Medicare fee-for-service (FFS) beneficiaries with an acute, nonspecific onset of LBP who had an initial provider visit at a participating practice and agreed to participate.
- A rigorous evaluation was not possible due to the inability to replicate the inclusion criteria using claims data and significant potential selection bias, limiting the ability to create a valid matched comparison group. Estimating impacts over all those eligible for the program would have yielded unbiased estimates, but would be unlikely to detect even very large true program impacts, due to the low participation rate and sample size.

Awardee leaders hypothesized that an evidence-based care delivery model that emphasized guideline adherence, access to care, educating patients, shared decision making, and care coordination would reduce inappropriate use of imaging, injections, and surgery. This in turn would reduce costs of care, improve patients' outcomes, and improve patients' satisfaction. Specifically, the awardee anticipated a reduction in imaging, surgery, and injection rates of 20, 20, and 25 percent, respectively; a reduction in Medicare spending per beneficiary per month of 3.7 percent; clinically significant improvement in patient-reported function and general health outcomes in 90 percent of patients; and patients' satisfaction scores 5 percentage points above national benchmarks for 90 percent of patients. During the program, the awardee dropped reduced injection and surgery rates as target outcomes because of their low incidence.

Table 1. Program characteristics at a glance

Program characteristic	Description
Purpose	The program aimed to encourage conservative, evidence-based treatments such as avoiding bed rest, managing pain, exercise, physical therapy, and education for patients with acute, nonspecific LBP and reduce unnecessary spinal imaging, injections, and surgery.
Major innovation	The awardee implemented an LBP checklist to function as a decision support tool for participating primary and specialty care providers and to prompt providers to follow an evidence-based treatment protocol for all patients presenting with acute, nonspecific LBP.
Program components	• Evidence-based treatment guidelines • Primary and specialty care coordination • Shared decision making
Target population	The program sought to engage Medicare FFS, Medicare Advantage, or Medicaid patients 18 years or older making their first visit to a BBC participating provider for acute, nonspecific LBP.
Participating providers	36 clinical sites participated in the BBC program.
Total enrollment	The awardee enrolled 1,472 beneficiaries, representing 9 percent of its original enrollment goal.
Level of engagement	Of the 1,472 enrollees, nurse care managers could reach 73 percent for the initial Day 2 call and 62 percent for the Week 2 follow-up call. Further, they reached 28 and 24 percent of enrollees for calls in Months 3 and 6, respectively. They could reach only 18 percent for the last intended call in Month 12.
Theory of change or theory of action	By focusing on evidence-based guidelines and noninvasive treatments, the program aimed to reduce rates of imaging, injections, and surgery, which in turn were expected to lower spending levels. By focusing on educating patients, avoiding bed rest, managing pain, exercise, and physical therapy, the program aimed to improve patients' functional status and general health outcomes. By focusing on shared decision making, timely access to care, and care coordination, the program sought to increase patients' satisfaction with care.
Award amount	$6,034,888
Effective launch date	The awardee began enrolling patients in February 2015.
Program setting	The awardee implemented the program in outpatient primary care and specialty practices.
Market area	Rural, urban, and suburban
Market location	Five-county region in central North Carolina
Target outcomes	• Decreased imaging, surgery, and injection rates by 20, 20, and 25 percent, respectively • Reduced spending levels among participants by 3.7 percent • Improved patient-reported physical function and general health by 90 percent • Increased patients' satisfaction by 5 percentage points above national benchmarks for 90 percent of participants
Payment model	The awardee encouraged adoption of the BBC model throughout UNC's health system and affiliates, where existing alternative payment models supported the value orientation of BBC.

Table 1 (*continued*)

Program characteristic	Description
Sustainability plans	After the program ended, the awardee continued to make BBC resources available to providers, including the EMR-integrated BBC checklist and patient education materials. The awardee also expected that some form of pain psychology and broader care management services (not for a dedicated LBP program) would be available to patients through other university resources.

BBC = Better Back Care; EMR = electronic medical record; LBP = low back pain; FFS = fee-for-service.

A rigorous impact evaluation of the BBC program was not possible, for reasons provided in Table 2. Therefore, this report describes only the demographic and health characteristics of Medicare FFS participants and does not present estimates of program impacts.

Table 2. Key features of descriptive analysis

Features	Description
Descriptive analysis	It was not possible to conduct a rigorous impact evaluation for the BBC program because of the ways the awardee identified and recruited participants. The evaluation could not replicate some of the BBC enrollment criteria (including physician referral, EMR diagnoses, and patients' history) using claims data. Providers might have referred patients whom they felt would benefit most from the program, and patients decided whether to participate. Enrolled beneficiaries were much healthier than eligible nonparticipating beneficiaries at participating practices who met the eligibility criteria assessable with claims and likely differed in ways not controllable with claims data. Only 5 percent of all claims-eligible beneficiaries enrolled in the program. Estimating impacts over all those eligible for the program would yield unbiased estimates, but would be unlikely to detect even very large true program impacts, due to the low participation rate and sample size. As a result, this report describes only the demographic and health characteristics of Medicare FFS participants before they enrolled in the program.
Intervention group used for descriptive analysis	The intervention group used for a descriptive analysis of baseline data included 350 of the 678 Medicare participants. It excluded 120 for whom the program could not find an LBP claim during the 14 days before program enrollment date, 105 with an LBP diagnosis claim in the six months preceding the initial BBC provider visit or with an exclusion diagnosis or treatment, and 40 from practices with 10 or fewer eligible enrollees during the program period. It also excluded 63 beneficiaries who did not meet the standard Medicare FFS eligibility criteria for this study.
Limitations	Due to the problems noted above, the evaluation cannot make inferences about impacts of this program on Medicare expenditures or service use.

BBC = Better Back Care; EMR = electronic medical record; FFS = fee-for-service; LBP = low back pain.

PROGRAM DESIGN AND ADAPTATION

The BBC model had three components: (1) evidence-based treatment guidelines, (2) primary and specialty care coordination, and (3) shared decision making.[1]

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Evidence-based treatment guidelines

The program designed an LBP checklist and a customized care manager decision support database that outlined evidence-based treatment guidelines for acute, nonspecific LBP. The decision support guided providers in using the most conservative—but appropriate—treatment approach. The awardee expected participating providers to use the electronic medical record (EMR)-embedded LBP checklist for any patient presenting with acute LBP. The checklist included a series of questions about LBP onset, LBP-related functional limitations, treatments to date, depression screening, red flags, and a basic physical examination. The checklist also required providers to identify the treatments and education offered to the patient. The EMR-embedded checklist automatically generated handouts delivered to patients describing best practices for LBP self-treatment. The handouts were also available through the exercise physiologist and the nurse care managers.

Program staff produced provider profile reports that documented practices' and providers' use of the LBP checklist and evidence-based LBP guidelines. Providers indicated that sharing of provider profiles and benchmarks consistent with the program's treatment guidelines encouraged constructive competition among providers, avoidance of inappropriate testing, and discussion on approaches to increase adoption of evidence-based practices. One physician leader indicated that the checklist-embedded decision support tools improved providers'—especially medical residents'—understanding of evidence-based conservative care, resulting in fewer orders for unnecessary imaging and injections.

Primary and specialty care coordination

Program staff, including a program director, program manager, and nurse care managers, identified a referral network of clinicians and spine specialists (for example, neurologists, rehabilitation and pain specialists, physical therapists, and exercise physiologists) to provide care consistent with the BBC model. After participants enrolled, nurse care managers contacted them within 48 hours of the first provider visit for LBP and again within two weeks. The care managers used the customized care manager decision support tool and information gathered through the Day 2 and Week 2 calls to determine the need and timeline for additional phone-based care coordination and support. This care coordination and support continued for as long as care managers deemed it necessary to treat an episode of acute LBP, but usually no longer than three months. Phone-based care coordination and support included facilitating access to exercise physiology classes; facilitating appointments with program-affiliated specialists and other professionals, such as the pain psychologist or physical therapist; and encouraging positive health behaviors, such as exercise.

Shared decision making

The BBC program engaged participants in shared decision-making during providers' visits and nurse care managers' phone consultations throughout their treatment episode to encourage the most conservative, yet appropriate, treatment approaches. The program offered participants

multiple opportunities for shared decision making beyond the initial visit via follow-up phone calls with care managers and referral to other LBP specialists, when participants sought and could benefit from additional discussions and support.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The awardee was effective in implementing BBC program components, though on a smaller scale than projected. The program leaders and staff felt that the program had several favorable effects, including: improved providers' adherence to evidence-based guidelines and best practices for acute LBP; increased patient satisfaction and education; reduced pain, partially through exercise physiologist referrals; and reduced unnecessary imaging and costs. However, the program encountered significant barriers to enrolling participants, including (1) considerably lower volume of patients with acute LBP than originally projected, (2) delays in implementing and activating the EMR checklist, and (3) missed physician opportunities in identifying and referring eligible patients due to limited engagement and competing demands. The awardee implemented several corrective strategies to increase enrollment, such as recruiting more practices, with a focus on practices with a high target patient volume; reinforcing providers' training and feedback; and supplementing physician referral with an EMR-based diagnosis report to identify participants. The awardee also found that the *International Classification of Diseases* (ICD) diagnosis codes were not sufficiently reliable or specific for distinguishing acute from chronic LBP, necessitating chart review and patient history to identify eligible participants.

Other implementation challenges included barriers in making the checklist operational across participating sites due to different EMR system configurations; staff turnover and leaves of absence that affected identifying and enrolling participants; and the reluctance of many providers to use the LBP checklist, educate patients, conduct depression screenings, or distribute patient education handouts. The awardee also encountered challenges with engaging participants, as described in the next section.

Enrolling and engaging participants

UNC originally identified participants through physician referrals using the LBP checklist. However, because physicians often missed referring potentially eligible patients to the program, UNC expanded its enrollment efforts to include nurse care managers identifying potential participants using an EMR-based daily diagnosis report, chart review, and patients' histories. Nurse care managers then provided a list of eligible patients to participating providers. During the Week 2 call, care managers asked patients to consent to participate in program data collection. If the patients consented, the care managers contacted them to participate in phone-based data collection on patient-reported health and patient satisfaction outcomes at 3, 6, and 12 months.

Engaging participants was a challenge, with nurse care managers being unable to reach about 27 percent of all participants for the initial Day 2 call and 38 percent for the Week 2 follow-up call. In addition, care managers could not reach 72 to 82 percent of participants for the Month 3 follow-up calls and beyond, resulting in large amounts of missing data on the patient-reported health and patients' satisfaction outcomes. Many participants also did not attend appointments with the BBC exercise physiologist or pain psychologist.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Study sample

The participants used in the analysis included 350 Medicare FFS beneficiaries who had a diagnosis of acute, nonspecific onset of LBP and an initial visit at one of the five high-volume BBC participating practices from February 2015 to August 2017 as identified in Medicare claims data *and* were reported as participating in the program by the awardee. This analytic sample represented roughly half (52 percent) of the 678 Medicare participants. Appendix A, Table A.1 shows the identification of the final participant sample from all 1,472 participants. The final sample excludes 581 participants who were not Medicare beneficiaries or were Medicare beneficiaries with invalid ID numbers in the awardee's finder file. It also excludes 213 participants with Medicaid-only coverage, leaving a total of 678 identifiable, participating Medicare beneficiaries. The analysis sample excludes 328 of these 678 Medicare participants--63 who did not have Medicare as their primary payer for both Part A and Part B; 120 for whom an LBP claim could not be found within 14 days preceding the enrollment date on the finder file; 95 with an LBP diagnosis claim in the six months before the initial BBC provider visit and 3 with spinal injection or surgery in this period; 7 with an exclusion diagnosis; and 40 from practices with 10 or fewer eligible enrollees during the program period. The resulting usable sample comprises 350 participants.

In general, the program participants looked much like the Medicare population overall, and were even slightly healthier, with an average hierarchical condition category (HCC) score of only 0.90. However, nearly 10 percent had major depressive or other psychiatric disorders.

Table 3. Baseline characteristics of Medicare FFS participants

Characteristics	Participants (N = 350)
Demographics	
Average age at enrollment, years	71
Age group, %	
Younger than 65	12
65 to 74	54
75 to 84	27
85 and older	7.4

Table 3 (*continued*)

Characteristics	Participants (N = 350)
Female, %	58
Race, %	
White	79
Black	17
Other	3.4
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	81
Disability insurance benefits	18
End-stage renal disease	0.6
Medicare and Medicaid dual status, %	
Medicare and Medicaid dual status	9.7
LBP diagnosis, %	
Primary diagnosis code only	59
Secondary diagnosis code only	33
Primary and secondary diagnosis codes	8.3
HCC score[a]	
Average HCC score	0.9
Chronic conditions, %	
Chronic obstructive pulmonary disease	7.7
Morbid obesity	7.1
Rheumatoid arthritis and inflammatory connective tissue disease	6.0
Major depressive, bipolar, and paranoid disorders	9.7
Congestive heart failure	8.6
Stroke, hemiplegia, or hemiparesis	2.0
LBP diagnosis in first 6 months of baseline year, %	12
Opioid or substance use disorder, %	2.9
Service use and spending during year before enrollment	
Average total payment all services, $ PBPM	569
Average number of ED visits (per 1,000 beneficiaries)	571
Any ED visit, %	25
Any spinal imaging, %	8.0

Sources: Mathematica's analysis of information from awardee's finder file through August 2017 and Medicare claims and enrollment data as of August 2018.

Note: The evaluation defined the baseline year as the 365 days before each beneficiary's enrollment date, which was the date of the beneficiary's first visit to a provider for the qualifying acute, nonspecific LBP episode. The evaluation measured all beneficiary characteristics during or as of the end of the baseline year. The statistics are weighted means, with participant weights proportional to the number of months during the baseline period that the participant was in Medicare FFS.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

ED = emergency department; FFS = fee-for-service; HCC = hierarchical condition category; LBP = lower back pain; PBPM = per beneficiary per month.

Challenges of estimating program impacts

Although the awardee provided a diagnosis code list for identifying patients with acute LBP in claims data, the awardee also relied on clinical judgement and other data unobservable in claims

in deciding which patients to enroll in the program. In addition, the awardee excluded patients with self-reported LBP duration longer than three months, an eligibility criterion that could not be applied using claims data.

Because these criteria for identifying patients could not be used to select a comparison group that matched the participants, this study explored two alternative approaches to estimating program effects. The first approach was to select the 350 Medicare FFS participants who met the eligibility criteria observable in Medicare claims and compare their outcomes to those of a matched comparison group. That approach was rejected because these participants were much healthier than the larger pool of 7,788 Medicare FFS beneficiaries who met the claims-based criteria for eligibility, 7,438 (95 percent) of whom did not participate in the program. Thus, the participants were a highly selected (by clinicians and self-selection) subset of patients meeting the eligibility criteria assessable in claims. The much lower prior service use and expenditures for participants than for eligible nonparticipants, as well as differences on other observable characteristics, suggest that comparing outcomes for these participants to a matched comparison group would yield severely biased (overly favorable) estimates of program impacts.

The second approach considered—the intent-to-treat approach—was also rejected. That approach would have avoided such selection bias problems by comparing average outcomes for all 7,788 treatment-eligible beneficiaries to those of a matched comparison group meeting the same criteria. However, this approach was not viable, because program effects would be concentrated solely in the 350 (4.5 percent) of eligibles who, according to the awardee's finder file, participated in the program and received the intervention. Thus, even if impacts on those who actually received the intervention were as large as 20 percent of the mean outcome, this implies that impacts measured over the full set of treatment-eligible beneficiaries would be less than 1 percent (4.5 percent of 20 percent = 0.9 percent). Finding a statistically significant estimate in the sample when the true effect is that small would be highly unlikely unless the number of treatment-eligible beneficiaries were much larger. Thus, an impact analysis would be futile and was not conducted.

CONCLUSION

Avoiding unnecessary imaging for patients with acute LBP has been a nationally recognized health care priority, emphasized by the Centers for Medicare & Medicaid Services through a publicly reported quality measure and the National Committee for Quality Assurance through its Healthcare Effectiveness Data and Information Set quality measure. Reducing unnecessary imaging among patients with LBP was the program's primary goal.

For the reasons discussed in the preceding section, the analysis was unable to determine whether the program achieved that objective. Imaging rates among Medicare FFS participants during the program period were substantially lower than the imaging rates of all treatment-eligible Medicare FFS beneficaries treated by the same five high-volume physicians during the pre-implementation period (data not shown). However, the severe selection bias made it impossible

to determine how much of this difference, if any, could be attributed to the BBC program. Furthermore, secular declines might have accounted for the lower rates for participants—imaging rates among even the nonparticipating eligible beneficiaries during the program period were markedly lower than those for the pre-implementation eligible beneficiaries. Similarly, the evaluation could not conduct a credible analysis of program impacts on hospitalizations, ED visits, or total expenditures because the enrolled Medicare FFS patients were far healthier than Medicare FFS patients who were eligible but did not enroll. This selection was due in part to the use of eligibility criteria not observable in claims data, provider selection of patients, and patient self-selection in deciding whether to enroll when invited. The very low participation rate among those meeting eligibility criteria assessable in claims precluded the use of an intent-to-treat approach using all eligible patients.

Although no conclusions can be drawn about program impacts, the program encountered several challenges that likely limited its effectiveness. These challenges included barriers in making the checklist operational across participating sites due to different EMR system configurations; staff turnover and leaves of absence that affected identifying and enrolling participants; and the reluctance of many providers to use the LBP checklist, educate patients, conduct depression screenings, or distribute patient education handouts. The awardee also encountered challenges with engaging participants, with many being unreachable by nurse case managers and not appearing for scheduled appointments.

PROGRAM SUSTAINABILITY

UNC discontinued the BBC program in October 2017, 10 months before its award ended in August 2018. Before discontinuing the program, the awardee transitioned program participants back to their primary care providers. Although BBC no longer exists formally, providers still have access to some program resources. UNC developed an Expected Practice Tool for LBP, based on BBC principles, designed to improve quality of care for patients with acute LBP. The awardee shared the tool across the health system, including with clinically integrated providers and networks. In addition, the awardee continued to make available other resources from the program, including the BBC checklist that integrates into university provider groups' EMRs, and patient educational materials that providers can print from the EMR and distribute to patients. The awardee also expects that psychology and care management services resembling those offered through BBC will be available to patients with acute LBP through other university resources.

UNC proposed payment model

UNC originally proposed developing an episode-based bundled payment to pay for the BBC program. However, the awardee abandoned this approach in the third program year after realizing it would not be able to calculate payment amounts given challenges accessing needed data. Instead, the awardee expected that its health system could pay for most features of the program through existing alternative payment models that the awardee believed already supported the value orientation of BBC.

Before the award ended, UNC had abandoned the episode-based payment model it had originally proposed to pay for BBC. It did so after realizing the model could not calculate viable, accurate payment amounts given challenges accessing needed data. Instead, the awardee focused on strategies to sustain elements of BBC that would not require separate external funding. For example, the awardee expected that existing funding streams could cover the costs of using the Expected Practice Tool for LBP throughout its health system. The awardee also filed a recommendation with the ICD-10 coding and maintenance committee (which advises coding changes and enhancements to Medicare's Merit-Based Incentive Payment System) to add ICD codes that differentiate between acute and chronic LBP.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

The descriptive analysis was conducted on 350 program participants, who comprised 24 percent of the 1,472 total program participants, and 52 percent of the 678 Medicare FFS participants (Table A.1). The program defined participants as having a qualifying visit to a participating Better Back Care (BBC) provider for acute, nonspecific low back pain (LBP) from February 2015 to August 2017, as reported in the awardee's finder file. The final analysis sample excluded 581 participants who were not Medicare beneficiaries or were Medicare beneficiaries with invalid beneficiary ID numbers in the awardee's finder file. It also excludes 213 participants with Medicaid-only coverage, leaving a total of 678 identifiable, participating Medicare beneficiaries. The analysis sample excludes 328 of these 678 Medicare participants—63 who did not have Medicare as their primary payer for both Part A and Part B; 120 for whom an LBP claim could not be found within 14 days preceding the enrollment date on the finder file; 95 with an LBP diagnosis claim in the six months before the initial BBC provider visit and 3 with spinal injection or surgery in this period; 7 with an exclusion diagnosis; and 40 from practices with 10 or fewer eligible enrollees during the program period. The resulting analysis sample comprises 350 participants.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants excluded from analytic sample	Number of participants remaining for analytic sample
Total program participants		**1,472**
Not enrolled in Medicare or could not be identified in Medicare enrollment files	581	891
Medicaid-only coverage	213	678
Did not meet standard Medicare FFS inclusion criteria	63	615
Other study exclusions:		
Had no qualifying LBP claim in 14 days before enrollment date	120	495
Had LBP claim in 6 months before qualifying LBP claim	95	400
Had spinal injection or surgery in 6 months before qualifying LBP claim	3	397
Had cancer, fracture, high-impact trauma, significant neurologic impairment, or pregnancy during or 6 months before qualifying LBP visit	7	390
Enrolled in practice with fewer than 10 claims-eligible participants	40	350
Final analytic sample		**350**

Sources: Mathematica's analysis of information from the awardee's finder file through August 2017 and Medicare claims and enrollment data as of August 2018.

FFS = fee-for-service; LBP = low back pain; UNC = University of North Carolina at Chapel Hill.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Ventura County Health Care

September 2020

Nick Theobald, Ed Kako, Eric Verhulst, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

VENTURA COUNTY HEALTH CARE

Ventura County Health Care, a public health agency based in Ventura, California, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Chronic Obstructive Pulmonary Disease (COPD) Access to Community Health (CATCH) program. The program aimed to provide home-based care for patients in Ventura County diagnosed with COPD. The program's innovations included providing care for COPD patients in the community setting, thereby reducing the frequency and severity of clinical exacerbations and related costly hospitalizations. In addition, CATCH established an incentive program that encouraged primary care providers (PCPs) to implement the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines for the care of patients with COPD. To implement the guidelines, PCPs received (1) training to more accurately diagnose COPD in one of four stages of disease severity and (2) evidence-based strategies for treatment that aligned with each stage of the disease. Participants in the CATCH program also received care and case management services from the program's registered nurses and registered respiratory therapists, both of whom coordinated care with the patient's PCP and specialist.

Important issues for understanding the evaluation

- The CATCH program trained willing PCPs in Ventura County to implement GOLD guidelines for treating patients diagnosed with COPD. The guidelines provide diagnosis rules for assigning patients into one of four stages of disease severity and provides strategies for treatment aligned with each stage of the disease.
- The CATCH programed enrolled both people diagnosed with COPD as well as those in risk of developing COPD. The program enrolled 2,040 participants, representing 82 percent of the its three-year enrollment target.
- Because it was not possible to use either Medicare or Medicaid claims to identify a comparison group matching the disease stage or at-risk characteristics of participants, this study does not present impact estimates.

The awardee hypothesized that increased education and training of providers and participants would lead to reduced COPD-related exacerbations and thus fewer expensive outcomes, including emergency department (ED) visits and hospital stays. By implementing the GOLD guidelines, COPD patients in Ventura would receive more accurate diagnoses through spirometry testing that categorizes COPD into four stages. With these more accurate diagnoses, Ventura could better align care to the patient's clinical needs. In addition, the program provided participants with case management services, which included home visits to assess risk factors in their homes and provided smoking cessation services. Table 1 summarizes the program's key characteristics.

Ventura proposed paying for the CATCH program through a pay-for-performance model. It combined a discounted bundled payment for treating COPD patients with an incentive payment for PCPs who followed evidence-based clinical guidelines for treating COPD. PCPs who completed the program's certification course were eligible for an incentive payment every six

months. The level of payment depended on the extent to which PCPs documented in the COPD PowerForm tool embedded in the electronic health record (EHR) system that they followed the clinical guidelines to treat patients with COPD.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Ventura County Health Care implemented the CATCH program to improve the quality of care for Ventura County residents with COPD by enhancing providers' and patients' awareness of the GOLD guidelines and expanding access to health care and resources for patients with COPD. The program also provided services, such as nicotine replacement therapy, for patients identified as at risk of developing COPD. The awardee expected these improvements to decrease the incidence of avoidable hospitalizations, ED visits, and visits to PCPs.
Major innovation	Although the GOLD guidelines are the accepted standard of care for patients with COPD, many health care providers are unaware of them. The goal of CATCH was to educate providers to follow the best practices available, using spirometry tests to determine COPD staging in addition to assessing symptoms to determine the level of care to provide COPD patients. The innovation also led to creating a pulmonary rehabilitation clinic in Ventura County (the CATCH clinic), which the awardee hoped to sustain as a new service after the cooperative agreement ended.
Program components	• Provider training • Patient and family engagement
Target population	The program sought to engage Medicare and Medicaid beneficiaries in Ventura County with COPD and those at risk of COPD as a result of exposure to smoke and second-hand smoke in Ventura County. The program also sought to train up to 185 PCPs in the GOLD standard of care.
Total enrollment	Ventura enrolled 2,040 participants, representing 82 percent of the original enrollment goal. By the end of the second program year, approximately 138 PCPs had received GOLD certification with the majority being within Ventura County Health Care.
Theory of change or theory of action	The awardee expected the CATCH program to lead to better COPD management by teaching the GOLD guidelines to PCPs, nurses, and staff in family clinics. It also expected the program to lead to improved outcomes in pulmonary function and quality of life. Finally, Ventura County Health Care expected the CATCH program to reduce hospitalizations, ED visits, and visits to PCPs, resulting in cost savings to CMS.
Award amount	$6,128,059
Effective launch date	The awardee began enrolling beneficiaries in January 2015, five months after award.
Program settings	Patients' homes and family clinics
Market area	Ventura County, California
Target outcomes	• Reduced health care costs from reduced hospitalizations, ED visits, and PCP visits • Increased access to health care and community-based resources • Reduced COPD exacerbations • Improved quality of life
Payment model	Ventura County Health Care's CATCH payment model combined a discounted bundled payment for treating COPD patients with an incentive payment for PCPs to encourage them to follow evidence-based COPD clinical guidelines. The payment model, however, was not implemented.

Table 1 (*continued*)

Program characteristics	Description
Sustainability plans	After its award ended on August 31, 2017, Ventura County Health Care did not continue the CATCH program. Instead, the awardee reported that the CATCH program led to a number of new initiatives to address issues and illnesses identified during the implementation of the program. The awardee also continues to expand COPD information exchange functionality through its information technology system.

CATCH = Chronic Obstructive Pulmonary Disease Access to Community Health; CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disease; ED = emergency department; GOLD = Global Initiative for COPD; PCP = primary care provider.

It was not possible to conduct a rigorous impact evaluation of the GOLD program because of how the awardee identified and recruited participants into the program, as well as the lack of available Medicaid data of acceptable quality covering the program period. As a result, this report describes only the demographic and health characteristics of Medicare FFS participants and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis.

Table 2. Key features of the descriptive analysis

Features	Description
Descriptive analysis	Due to the small number of Medicare FFS beneficiaries who had claims with a COPD diagnosis, the limitations of the diagnostic data, and issues with the timeliness and reliability of California Medicaid encounter data, it was not possible to conduct a rigorous impact evaluation. Further, we did not have usable provider billing identification numbers to draw an intent-to-treat sample of beneficiaries for use as an alternative estimation approach. This report presents only the demographic, health, and service use characteristics for Medicare FFS participants during the year before they enrolled in the program.
Intervention group for descriptive analysis	The intervention group for the descriptive analysis relied on the 587 Medicare FFS participants, which includes Medicare only and Medicare and Medicaid dual enrollees (out of 2,040 total participants). The 1,453 excluded beneficiaries included 1,152 participants identified as Medicaid only or whose insurance status could not be determined and 301 Medicare participants who did not meet study eligibility criteria of being in Medicare Parts A and B FFS for at least three months before enrollment, Medicare as primary payer, and having a valid date of enrollment.
Limitations	Due to the problems noted above, the evaluation cannot make inferences about the impact of this program on Medicare costs or other program outcomes.

COPD = chronic obstructive pulmonary disease; FFS = fee-for-service.

PROGRAM DESIGN AND ADAPTATION

The CATCH service delivery model included two components: (1) provider training and (2) patient and family engagement.

Provider training

Ventura County Health sought to train up to 185 PCPs to refer eligible beneficiaries to the CATCH program and to integrate the GOLD guidelines into their clinical decisions when treating patients with COPD. In Year 1, CATCH staff focused on training providers within Ventura County Health Care and during Year 2 began to engage providers outside of the Ventura County Health Care Agency system. By the end of the second program year, approximately 138 PCPs had received GOLD certification with the majority being within Ventura County Health Care. To increase referrals in Year 3, Ventura County Health Care updated the Cerner EHR system within their health care system to integrate a pop-up referral option to CATCH as PCPs entered eligible diagnoses into the system. The awardee also trained PCPs to use spirometers to diagnose COPD by disease stage and then educated them on the proper treatment for each stage. The awardee provided training to nurses and CATCH staff on medication regimens and the use of respiratory equipment. In the third year of the program, the awardee began providing nicotine replacement therapy (NRT) and training staff to provide NRT to participants.

Engaging patients and families

To engage patients and families, nurses visited the homes of patients diagnosed with COPD to help them with their care regimes and address any environmental issues related to COPD. After the initial home visit, CATCH maintained contact with participants via text messaging at least once a month. Participants with more severe COPD or who had additional needs could call CATCH staff at any time or could visit the CATCH clinic for care and support when necessary. For patients diagnosed with COPD and those at risk of developing the disease, staff encouraged participation in smoking cessation programs and provided free NRT. As Ventura County Health Care began closing down the program at the end of the cooperative agreement, it connected participants with high needs to other case management services available in the county.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Ventura County Health Care began enrolling beneficiaries in January 2015, five months after award. The awardee reported that it enrolled 2,040 participants by the end of the program in August 2017, achieving 82 percent of its enrollment goal. Ventura County Health Care explained that two factors helped it come close to achieving its enrollment goal: (1) providing spirometers and GOLD guidelines training to providers and (2) providing NRT to at-risk COPD patients. Offering spirometers to providers was an incentive for them to join the CATCH program, which led to more referrals and thereby increased enrollment. Spirometers cost $1,500 to $2,000, so covering the cost of the devices ensured that providers willing to provide pulmonary function tests (PFTs) could do so without having to supply their own funding. CATCH staff also found providing NRT to patients was an effective recruitment tool, especially for the at-risk population. By the third year of the program, many stakeholders in Ventura County had heard of the CATCH

program, which made it easier for Ventura County Health Care to motivate additional providers to participate in the training on GOLD guidelines and enroll beneficiaries into CATCH.

As evidence of implementation effectiveness, the awardee reported a 17 percent increase in participation in smoking cessation programs, from 37 percent of known smokers attending the clinic before CATCH to 43 percent afterward. Over the same period, the percentage of participants receiving inhaled bronchodilator therapy increased 85 percent, from 18 percent of all participants to 33 percent. Bronchodilator therapy is the crucial, first line of treatment for COPD patients.

CATCH staff identified two facilitators of service delivery effectiveness: (1) completing the CATCH clinic and (2) the program's adaptability and the process for customizing the care plan for each participant. Completing the CATCH clinic in the second program year created a space for participants to receive specialized services, such as administering six-minute walking tests or arranging for pulmonologist visits for patients with late-stage COPD. The clinic enabled staff to accelerate the treatment of advanced COPD patients and to provide pulmonary rehabilitation, nutrition education, and counseling for participants with more severe stages of COPD. The process for customizing a care plan for each participant reflected the program's adaptability. Each week CATCH staff met to discuss each new participant and his or her plan of care. The plan focused not only on medication and services directly related to COPD, but also—and more comprehensively—on each individual's needs, including behavioral health and social needs such as food and housing.

Implications of program implementation for achieving program goals

- A self-reported increase in the percentage of patients engaged in program services, such as attending smoking cessation programs and receiving inhaled bronchodilator therapy (a first line of treatment for COPD), might have increased the likelihood of achieving program goals.
- However, staff and resource shortages, including the number of spirometers available for conducting a pulmonary function test (an assessment critical to the proper diagnosis of COPD) might have limited program effectiveness.

The awardee, however, faced several challenges implementing its program. Bureaucratic hurdles to engaging providers outside the Ventura County Health Care Agency system created challenges for enlisting providers necessary to reaching enrollment targets. Beginning in Year 2, the awardee engaged independent physician associations and provider groups outside the system to refer participants to the program. This was necessary to reach prospective participants outside the county system (and meet the enrollment target), but it was more complicated to share patients' information outside the system. The awardee had to meet data-sharing requirements for each partner organization before it could begin enrolling their beneficiaries in the CATCH program.

Staff and resource shortages affected how Ventura County Health Care Agency implemented the program. Initial shortages in nurses hampered the awardee's ambitious vision of providing each participant with a home visit. Home visits required a nurse to travel across a large county to each participant's home for a visit that could take more than an hour. Often there were not enough nurses to do so in a timely fashion.

CATCH staff also indicated that shortages in the number of spirometers available to providers hindered the ability to provide pulmonary function tests, which were necessary to accurately diagnose COPD. Ventura County Health Care reordered 70 spirometers beyond the initial distribution because it ran out of units to give providers. This temporary shortage might have led to a decrease in the rate of documented spirometry, because CATCH staff continued to enroll beneficiaries even though there were fewer resources to perform the PFTs. Following the distribution of the additional spirometers, the PFT rate at enrollment increased to 81 percent.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Characteristics of program participants

The intervention group for the descriptive analysis relied on the 587 Medicare FFS participants, which includes Medicare only and Medicare and Medicaid dual enrollees (out of 2,040 total participants). The 1,453 excluded beneficiaries included 1,152 participants identified as Medicaid only or whose insurance status could not be determined and 301 Medicare participants who did not meet study eligibility criteria of being in Medicare Parts A and B FFS for at least three months before enrollment, Medicare as primary payer, and having a valid date of enrollment. Appendix A, Table A.1 describes the identification of the sample used for the descriptive analysis.

Table 3 displays descriptive characteristics for the 587 Medicare FFS participants. The average age of the Medicare FFS participants was 64, and about half were male (47 percent). Three-quarters of the participants were White. Hispanics represented the next-largest ethnic group at 14 percent. Of the FFS participants, 78 percent were Medicare and Medicaid dual enrollees—more than four times the share in the general Medicare FFS population (18 percent)—which indicates a high level of economic need. CATCH participants also had substantial health care needs. The FFS participants were about twice as likely as their counterparts in the general Medicare population to have become eligible for Medicare because of a disability (63 versus 34 percent, respectively). Furthermore, the average hierarchical condition category (HCC) risk score for CATCH Medicare FFS beneficiaries was 60 percent above the average score for Medicare FFS beneficiaries nationwide. Nearly one-quarter of Medicare FFS participants were not assigned a COPD stage at enrollment, and among those who were, the most common stage was 0, reflecting individuals who did not have COPD but were at risk of developing it.

Table 3. Baseline characteristics of Medicare FFS participants

Characteristic	Medicare FFS participants (N = 587)
Demographics	
Age at enrollment, years	64
Male, %	47
White, %	76
Black, %	4
American Indian, Alaska Native, Asian/Pacific Island American, or other, %	6
Hispanic, %	14
Unknown, %	< 1
Original reason for Medicare eligibility, %	
Old age and survivor's insurance	36
Disability insurance benefits	63
ESRD[a]	0.35
Medicare–Medicaid dual status, %	
Not dually eligible	22
Dually eligible	78
HCC score[b]	
Mean	1.6
Chronic conditions, %	
COPD	42
Stage of COPD, %	
No stage	24
0 (at risk)	29
1 (very mild)	4
2 (moderate)	19
3 (severe)	17
4 (very severe)	7

Sources: Mathematica's analysis of information from awardee's finder file and Medicare claims as of August 31, 2017.

Notes: The evaluation defined the baseline year as the 365 days before each beneficiary's enrollment date. It defined the enrollment date as the date on which the beneficiary consented to participate in the program. It measured all beneficiaries' characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

[a] Identified in the last month of each beneficiary's baseline year.

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms

COPD = chronic obstructive pulmonary disease; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category.

Table 4 displays expenditure and service use by COPD stage for the year before Medicare FFS beneficiaries entered the CATCH program. Average Medicare expenditures were $1,358 per beneficiary per month in the baseline year. Participants with Stage 0 or Stages 1 or 2 showed similar levels of expenditures just above $1,000, and those with Stages 3 or 4 showed

substantially higher levels of expenditures at $1,746 per beneficiary per month. Overall, there were 551 hospitalizations per 1,000 beneficiaries, with Stage 0 participants showing the lowest level of hospitalizations at 303 per 1,000 beneficiaries (similar to the national average for all Medicare beneficiaries), and those classified with Stages 3 and 4 showing the highest at 788. For ED and PCP visit measures, those with mild or moderate COPD (Stages 1 or 2) had the lowest rates of service use, and those at risk of developing COPD (Stage 0) showed similar rates of service use as the highest COPD severity (Stages 3 or 4) participants.

Table 4. Baseline expenditures and service use of Medicare FFS characteristics, by COPD stage

Measure	All (N = 587)	COPD stage		
		Stage 0 (N = 169)	Stages 1 and 2 (N = 133)	Stages 3 and 4 (N = 142)
Total expenditure ($ PBPM)	1,358	1,021	1,059	1,746
Hospitalizations (per 1,000 beneficiaries)	551	303	429	788
ED visits (per 1,000 beneficiaries)	1,300	1,331	1,011	1,334
PCP visits in ambulatory setting (per 1,000 beneficiaries)	6,271	6,258	5,966	6,497

COPD = chronic obstructive pulmonary disease; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; PCP = primary care provider. The 143 beneficiaries who did not have a COPD stage were excluded from this table.

Challenges of estimating program impacts

A rigorous impact evaluation of this program was not possible for three reasons. First, due to concerns about the quality of California Medicaid encounter data and the inability to reliably match CATCH participants with these data, it was not possible to measure outcomes for Medicaid beneficiaries nor to create a comparison group of Medicaid beneficiaries. Second, only 335 of the 587 Medicare FFS participants had a COPD diagnosis in their claims history. The other 252 beneficiaries were diagnosed as being at risk for developing COPD, which claims data cannot capture. Further, the stage of COPD is not available in claims; this is a critical factor in drawing a similar comparison group. Because a comparison group could not be identified, this report does not present impact estimates. Using an intent-to-treat approach by comparing outcomes for all Medicare FFS beneficiaries in the catchment area who received care from participating providers and met the eligibility criteria to a matched comparison group would address this bias. However, that approach was not feasible because we did not have usable provider billing identification numbers to draw an intent-to-treat sample of beneficiaries.

CONCLUSIONS

Ventura County Health Care reported being mostly successful in implementing the CATCH program. The program faced challenges in coordinating with providers outside the Ventura County Health Care system that slowed enrollment. Shortages in the number of available spirometers hindered the ability to make COPD diagnoses at intake, and initial shortages in

nurses hampered the awardee's ambitious vision of providing each participant with a home visit shortly after enrollment. Even so, CATCH achieved 86 percent of its target enrollment goal, and the awardee reported growth in participation in its smoking cessation program and increases in the numbers of participants receiving inhaled bronchodilator therapy. Due to the small number of Medicare FFS beneficiaries and the lack of equivalent diagnostic information on claims for potential comparison cases, it was not possible to select a similar comparison group to conduct a rigorous impact evaluation.

PROGRAM SUSTAINABILITY

After its award ended in August 2017, Ventura County Health Care did not continue the CATCH program. Instead, the awardee reported that the program led to a number of new initiatives—such as a falls prevention program, which addresses concerns raised during CATCH meetings about falls, and a focus on billing Medicare for new COPD-related services, such as the Breathe Technologies open ventilator and distance monitoring and counseling. The awardee also continued to expand COPD information exchange functionality through its information technology system.

Ventura County Health Care originally proposed two types of payment to pay for CATCH: a discounted bundled payment for treating COPD patients for primary care practices and an incentive payment for PCPs who followed evidence-based COPD clinical guidelines. Participating PCPs had to meet two requirements to receive incentive payments: (1) complete a CATCH continuing medical education certification course and follow best practices (including the GOLD guidelines) and (2) use the PowerForm tool in the EHR. The COPD PowerForm was available in the Cerner EHR system that all providers could access (and linked to the bundled service). The awardee, however, did not pursue any payers to fund the payment model other than the Centers for Medicare & Medicaid Services, which declined to move forward with the bundled payment component. Without any payers on board, the physician incentive payments—for which Ventura County Health Care had used award funding—also ended after the award.

Ventura's proposed payment model

Ventura proposed paying for the CATCH program through a pay-for-performance payment model, which combined a discounted bundled payment for treating COPD patients with an incentive payment for PCPs who follow evidence-based clinical guidelines for treating COPD. PCPs who completed the program's certification course were eligible for the incentive payments every six months. The level of payment depended on the extent to which PCPs documented in the COPD PowerForm tool embedded in the EHR that they followed the clinical guidelines to treat patients with COPD.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total Medicare beneficiaries in awardee's finder file as of August 31, 2017		**888**
Missing enrollment date	1	887
Did not meet study's standard claims-based inclusion criteria		
Not enrolled in both Part A and B	138	749
Enrolled in Medicare Advantage	148	601
Medicare not primary payer	4	597
Fewer than 90 days of claims history before enrollment	10	587
Final Medicare FFS beneficiaries in descriptive analysis		**587**

Source: Mathematica's analysis of information from awardee's finder file and Medicare claims as of August 31, 2017.
FFS = fee-for-service.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: The Village Center for Care

September 2020

Cordon Newhart, Jim Derzon, Ellen Wilson, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

THE VILLAGE CENTER FOR CARE

The Village Center for Care (VillageCare), a community-based nonprofit organization in New York City, received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create and support the implementation of the Rango program. The goal of the program was to provide support to HIV-positive participants via an interactive website (Rango.net) and a mobile application. Rango sought to help adults ages 18 and older who were HIV-positive living in New York City and its surrounding area maintain their health by taking their antiretroviral medications every day as prescribed. The program launched in April 2015 (eight months after award) and the intervention period funded by HCIA R2 ended in February 2018. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- Improving treatment adherence was a major goal of the program, yet the awardee's data indicate that 82 percent of participants were adherent to treatment at enrollment. Rango could affect adherence only among the relatively small proportion of participants who were nonadherent at enrollment and were actively engaged in the program.
- There was likely some self-selection bias because the participants who already effectively managing their own care were the ones who most actively engaged in the program, whereas those who most needed support were less likely to enroll and use the resource.
- Program leaders and partners reported that it would take more than a year or two for behavioral modifications to manifest as changes in health outcomes, hospitalizations, and health care costs. As a result, the full impact of Rango on service use and expenditures might not be observed due to the short (26 months on average) follow-up period.

The Rango website contained a variety of features to support disease self-management, including a community forum on which participants could post comments and interact with one another; articles with useful information about HIV care and treatment; and treatment, medication, and appointment reminders. Participants enrolled in the program for 12 months. After 12 months, they could continue using most of Rango's mobile application services, but they would no longer receive the $35 to $40 monthly participation incentive or appointment and medication reminders via text message (although they could receive reminders via the Rango application). The goals of the program were to (1) increase participants' retention in HIV care; (2) increase treatment adherence; (3) increase participants' time in first-line treatment (that is, least burdensome and least costly); and (4) reduce costly hospitalizations and outpatient services associated with treatment failure.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Improve adherence to HIV treatment through the use of an integrated mobile platform and application.
Major innovation	Using an innovative technological approach to improve adherence to HIV treatment consisting of an integrated mobile platform and a mobile application provided participants a community forum on which they could post comments and interact with one another; articles with useful information about HIV care; and treatment, medication, and appointment reminders.
Program components	• Program liaisons to aid in recruitment and facilitate enrollment • Health coaches to develop Rango content and respond to participants' questions • Health IT via a mobile platform and application with educational, motivational, and reminder features
Target population	New York City residents and surrounding area residents ages 18 and older who were diagnosed with and prescribed medication for HIV and were covered by Medicaid, Medicare, or both (dually eligible)
Participating providers	Eight MCOs provided patient referrals for the duration of the cooperative agreement and the awardee added noncontractual referral partners in subsequent years.
Total enrollment	The awardee enrolled 4,367 participants.
Level of engagement	Nearly 90 percent of enrollees logged into the Rango platform at least once. As of November 2017, 4,367 participants had signed into Rango 277,700 times (an average of 63 times per participant); posted 32,434 messages (an average of more than 7 messages per participant); accessed articles 14,973 times (an average of more than 3 times per participant); and received 1.1 million medication reminders (an average of 251 reminders per participant).
Theory of change or theory of action	Participants' use of electronic self-care tools will improve their adherence to HIV treatment and their engagement in and satisfaction with care. VillageCare anticipated that Rango would ultimately reduce the costs associated with treatment failure and eliminate the need for more burdensome and expensive therapies.
Award amount	$7,983,297
Effective launch date	The Phase 1 Rango platform became available in April 2015. The Phase 2 platform, implemented in the second program year, dropped underused features (for example, virtual support groups and connecting to peer mentors) and added others (for example, live chat with health coaches, a searchable social services database, and enhanced communication features on the home page).
Program settings	• Recruitment occurred at partnering community-based organizations, MCO sites, and primary care provider locations. • Services were delivered through the Rango platform.
Market location	New York City and surrounding areas—including the Bronx, Brooklyn, Manhattan, Queens, and Staten Island, as well as Nassau, Suffolk, and Westchester counties
Target outcomes	• Improved adherence to treatment • Improved participant engagement • Improved participant satisfaction
Payment model	VillageCare's proposed payment model consisted of a PBPM payment to the program from an entity that was responsible for the health of its population, such as an ACO or MCO. The awardee leaders described two levels of payment, one intended to cover only the Rango software and another higher-level payment to cover the software plus VillageCare staff who ran the program, such as the enrollment staff and health coaches.

Table 1 (*continued*)

Program characteristics	Description
Sustainability plans	To sustain the Rango program, VillageCare worked to gain approval from two MCOs for the payment model. If Rango showed evidence of effectiveness and the cost was low enough, one MCO was interested in providing payment and expanding the program to its other members. VillageCare also tried to obtain grants from federal agencies and philanthropic foundations and potentially scale the program to other providers and populations. However, VillageCare terminated the program when its award ended in February 2018 because it did not have other funding sources to sustain the program.

ACO = accountable care organization; FFS = fee-for-service; IT = information technology; MCO = managed care organization; PBPM = per beneficiary per month.

Although 4,367 beneficiaries enrolled in the program, the impact analysis excluded those with incomplete Medicare FFS data or incomplete Medicaid data as well as those who did not meet program eligibility criteria. After making the sample exclusions explained in Table 2 and in Table A.1 of Appendix A, the impact analysis relied on the 420 Medicare and 2,532 Medicaid beneficiaries who enrolled in the Rango program from April 2015 through February 2017. The comparison group consisted of 1,946 Medicare and 8,873 Medicaid beneficiaries ages 18 and older with HIV living in New York City or the surrounding counties with similar demographic and health characteristics as those who were not enrolled in the program. Table 2 summarizes the key features of the impact evaluation.

Table 2. Key features of program evaluation

Features	Description
Evaluation design	The analysis relied on a difference-in-differences model that compared the change in outcomes among study beneficiaries after versus before enrollment relative to the change in outcomes over the same period among a matched comparison group.
Intervention group for evaluation	Of the 4,367 participants, 211 could not be linked to the Medicaid or Medicare data, 955 were Medicare beneficiaries (including dual eligibles) and 3,201 were Medicaid beneficiaries. Of the 955 Medicare beneficiaries, 428 were from the study because they were enrolled in Medicare Advantage, 100 were excluded because of not having sufficient Medicare FFS data and 7 were excluded because they did not meet program eligibility criteria or were missing a key matching variable. Of the 3,201 Medicaid participants, 468 were excluded because they did not meet program eligibility criteria and 201 were excluded because they did not have complete Medicaid data. After making these exclusions, the impact analysis included 420 Medicare FFS and 2,532 Medicaid beneficiaries who enrolled in the Rango program from April 2015 through February 2017.
Comparison group	The comparison group consisted of 1,946 Medicare and 8,873 Medicaid beneficiaries that met program eligibility criteria (ages 18 and older with HIV living in New York City or the surrounding counties) and had similar demographic and health characteristics as those who were not enrolled in the program
Limitations	If participants differed from eligible nonparticipants in ways not captured in Medicare administrative files and claims, the impact estimates might be biased. The low participation rate (about 23 percent) would have made it difficult to identify impacts if measured over all eligible beneficiaries.

FFS = fee for service.

PROGRAM DESIGN AND ADAPTATION

The VillageCare program's service delivery model had three components: (1) program liaisons, (2) health coaches, and (3) health information technology (health IT).[1]

Program liaisons

After providers referred potential participants, program liaisons screened them for eligibility and enrolled those who were qualified and interested into the Rango program. At first, liaisons conducted most enrollments by appointment. The program liaison walked participants through all of the paperwork (informed consent, data sharing authorization, and registration forms) and provided a demonstration of the Rango platform. This process proved too time consuming, so by the third year, program liaisons had started enrolling most participants at the referring organization (for example, in waiting rooms) instead of by appointment. The liaisons gave participants the enrollment packet to complete on their own. Program liaisons answered participants' questions and occasionally conducted a short demonstration of the platform or helped participants download the Rango application on their mobile devices, if needed. Participants could also schedule a full demonstration of the platform and application.

Health coaching

The health coaches developed content for Rango and engaged participants by hosting monthly wellness challenges, promoting events and discussions, and encouraging participants to think about their health. They also wrote monthly articles for the online library and interacted directly with participants through discussion boards, live chat sessions, email, and phone calls. Two health professionals staffed the health coaching positions: one was a licensed social worker and the other a dietician. Although program staff and leaders thought that participants' interactions with the health coaches played a key role in keeping participants engaged, it was not a highly used feature. Only 4 percent of participants held a live chat with a health coach via the website and 2 percent chatted with a health coach via the application.

Health IT

The integrated mobile platform (Rango.net) and the Rango mobile app were the principal components of the VillageCare intervention. Rango contained educational, motivational, and reminder features to provide support to HIV-positive participants in maintaining their health by taking their antiretroviral medications as prescribed. There were two planned phases to the Rango roll-out. The Phase 1 Rango launch included (1) text reminders to take and refill medications; (2) a medication tracker; (3) a user profile and avatar (user thumbnail photo); (4) an online community of other participants via discussion boards, friend requests, member searches, and private messaging; (5) contact forms to request customer service; (6) a library of self-help articles; (7) round-the-clock virtual support groups on a variety of topics; (8) telephone question-

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

and-answer sessions and messaging with health coaches; and (9) a connection to trained peer mentors.

After reviewing the first-year data on program use and participants' feedback, VillageCare rolled out a revised version of Rango (Phase 2) in the middle of the second year. Phase 2 dropped some underused features (such as the virtual support groups and connections to peer mentors) and added others (such as live chat with health coaches, a searchable social services database, and enhanced communication features on the home page). After the Phase 2 launch, VillageCare continued to make minor refinements to the platform in collaboration with its technology vendors.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

VillageCare developed the Rango platform and mobile-enabled website on time, with an ambitious schedule of launching seven months after the start of the cooperative agreement. However, due to a delay developing the mobile application, it was not available until halfway through the second year. Although program leaders planned and budgeted for two stages of development, they thought that a longer time frame for development and testing and more formative research might have improved the initial effectiveness of the intervention. The awardee spent time and money developing features that people did not use, and those resources might have been better used if the awardee had received more advance input from participants and providers.

Hiring and retaining staff as program liaisons was also a challenge that adversely affected enrollment and disrupted the smooth operation of the program throughout the life of the award. In Year 1, competition from other organizations for similarly qualified staff created difficulties in hiring for the program liaison positions. Internal conflicts between staff members in Year 2 and liaisons leaving in Year 3 due to concerns about being unemployed when the enrollment period ended impeded recruitment.

Finally, participants' challenging life circumstances and limited experience with technology were significant barriers to

Implications of program implementation for detecting impacts

- VillageCare continuously improved Rango, releasing a mobile app and Phase 2 version in the second year that was responsive to participants' needs and interests. If these increased the frequency with which participants used Rango and/or made Rango more effective, then the program might have stronger effects on participants who enrolled later in the program.
- Program administrators reported that some participants might not have been meaningfully engaged in Rango and were motivated to participate only by the financial incentive. Rango would be less likely to affect service use and spending among these participants.
- Participants who already effectively managing their HIV were reportedly most likely to engage in Rango. This meant that it would be unlikely for the program to have impacts because many participants had little opportunity to improve their self-care.

delivering effective services. HIV-positive patients in general tend to face multiple economic and social challenges, especially those who struggle to manage their care. In addition, the target population's lack of technical experience meant some participants struggled to use Rango and they required extensive customer support, particularly during the first two years of the program. This led to frustration among participants (if they could not get a timely answer to their technical questions), posed a burden for partner organizations (if participants kept coming back to them for help), and strained human resources at VillageCare.

ESTIMATING PROGRAM IMPACTS

Study sample

Although 4,367 beneficiaries enrolled in the program, the treatment group consisted of the 2,532 Medicaid and 420 Medicare treatment group beneficiaries who enrolled in the program from April 2015 through February 2017, had sufficient Medicaid or Medicare data to be included in the analysis, and met the claims-based eligibility criteria (that is, had a diagnosis of HIV and at least 18 and living in New York City or the surrounding areas). The comparison group included 1,946 Medicare and 8,873 Medicaid matched beneficiaries with similar demographic and health characteristics who met the same eligibility criteria, but were not enrolled in the program; comparison group members theoretically could have enrolled in the program, but most were likely unaware of it.[2]

Engaging participants

The theory of change advanced by the Rango program was that participants' use of electronic self-care tools would improve their adherence to HIV treatment directly through medication reminders and indirectly through social support. The awardee expected this combination of support to improve participants' engagement in effective health care self-management. However, Rango enrolled less than 2 percent of both Medicaid and Medicare beneficiaries who met the program's eligibility criteria (according to Medicaid and Medicare claims data), and those who did enroll were a highly motivated group. The vast majority of participants (82 percent) were adherent when they enrolled and reportedly already managed their disease well. Rango could meaningfully improve treatment adherence only among the 18 percent of participants (532) who were nonadherent at enrollment. Moreover, participants who already managed their disease well were also those who most actively engaged in Rango, whereas those who needed the most support were less likely to enroll and less likely to use the program if they did enroll. It might be that Rango will have an impact only on the relatively small proportion of participants who were nonadherent at enrollment and who were actively engaged in the program.

[2] More than 150,000 Medicare and Medicaid beneficiaries were eligible for the program. It is possible that the program recruited some comparison beneficiaries who chose not to participate; however, recruitment likely reached only a small proportion of eligible beneficiaries. Most beneficiaries contacted by the awardee enrolled; for example, the awardee's data suggested that 62 percent of referred beneficiaries who came in through its appointment system signed up for the program.

Although the program did not meaningfully engage a significant subset of participants, overall use of Rango was widespread and treatment adherence increased to 97 percent by program completion (as reported in the awardee's data). Although use of Rango varied among participants, overall use was high with 89 percent of enrollees logging into Rango at least once. The most widely used program features were receiving medication reminders (50 percent), posting messages in the community forum (37 percent), and accessing HIV health-related articles (33 percent). As of November 2017, the 4,367 participants had signed in to Rango a total of 277,700 times (an average of 63 times per participant); posted 32,434 messages (an average of 7.4 times per participant); accessed articles 14,973 times (an average of 3.4 times per participant); and received 1.1 million medication reminders (an average of 251 reminders per participant).

Characteristics of treatment and comparison group beneficiaries

Comparing treatment and comparison group characteristics at baseline confirmed that the two groups were well balanced on observable measures (see Appendix B), and the Medicare and Medicaid beneficiaries shared enough characteristics to be pooled for analysis (Table 3). Most participants (86 percent) were Medicaid beneficiaries; only 14 percent were Medicare beneficiaries. About 90 percent of the Medicare beneficiaries in the study were dually eligible for Medicaid. (To avoid double-counting, only the Medicare analysis included the dually eligible.) Most participants were ages 45 to 64, male, and of Black, Hispanic, or unknown race

The intervention enrolled only people with HIV, so the beneficiaries had high service use during the baseline period. In the 12-month baseline period, more than 40 percent had experienced an outpatient emergency department (ED) visit. Beneficiaries were at very high risk for increased service use, as the average beneficiary had risk scores that were much higher than 1.0 (the population average); specifically, Medicaid beneficiaries had an average Chronic Illness and Disability Payment System (CDPS) score of 4.7 while the average Medicare beneficiary had a hierarchical condition category (HCC) score of 1.7.

Table 3. Baseline characteristics of treatment and comparison group beneficiaries

	Medicaid beneficiaries		Medicare beneficiaries	
Measure	Treatment (N = 2,532)	Comparison (N = 8,873)	Treatment (N = 420)	Comparison (N = 1,946)
Demographics				
Age at enrollment, years	47	47	54	54
Age group, %				
18 to 44	32.8	35.0	13.6	17.7
45 to 54	44.0	38.0	40.0	31.0
55 to 64	23.0	26.0	31.0	33.0
65 or older	0.2	0.3	14.4	18.6
White, %	0.7	1.4	22.0	22.0
Black, %	2.6	4.5	57.0	57.0

Table 3 (*continued*)

Measure	Medicaid beneficiaries		Medicare beneficiaries	
	Treatment (N = 2,532)	Comparison (N = 8,873)	Treatment (N = 420)	Comparison (N = 1,946)
Hispanic/Latino, %	34.0	34.0	17.0	17.0
Asian, %	0.0	0.2	3.6	2.8
Other race, %	0.0	0.06	0.0	0.0
Unknown race, %	63.0	60.0	0.0	0.7
Gender: Male, %	63.0	62.0	67.0	67.0
Dually eligible for Medicare and Medicaid, %	n.a.	n.a.	91.9	92.7
Original reason for Medicare entitlement is disability, %	n.a.	n.a.	90.2	88.6
Percent of Medicaid beneficiaries in managed care	92.9	97.1	n.a	n.a
Health status				
CDPS score[a]	4.7	4.6	n.a.	n.a.
HCC score[b]	n.a.	n.a.	1.7	1.7
Service use and expenditures during the year before enrollment				
Any hospitalizations, %	24	24	21	21
Any outpatient ED visits, %	46	42	43	41
Total Medicare expenditures ($ PBPM)	n.a.	n.a.	1,391	1,347

Source: Mathematica's analysis of information from the awardee's program encounter database, and Medicare and Medicaid claims, encounter, and enrollment data from September 2012 through May 2018.

Notes: The baseline period covers the 12-month period before enrollment and periods are beneficiary specific. Enrollment ended February 28, 2017.

Expenditures could not be estimated for Medicaid beneficiaries because nearly all were in comprehensive managed care. Medicare beneficiaries were only included in the analysis while they were FFS (and not in managed care). The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare. In addition to the number of months enrolled in Medicaid or in FFS Medicare, the study weighted the statistics for comparison beneficiaries to reflect the number of times a comparison beneficiary was matched to a treatment beneficiary.

None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level, 2-tailed test.

Appendix B presents full balance results. Exact matching variables for the Medicare analysis include the quarter of enrollment for treatment beneficiaries and the pseudo-enrollment date for the comparison beneficiaries. Exact matching variables for the Medicaid analysis include the quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, and enrollment in comprehensive managed care.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending

[b] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

CDPS = Chronic Illness and Disability Payment System; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; n.a.= not applicable; PBPM = per beneficiary per month.

Analytic approach

The evaluation based the impact estimates on a difference-in-differences study design. This design measures program effects as the change in outcomes among study participants before versus after enrollment relative to the change in outcomes among a comparison group with similar characteristics over the same period. Assuming that external trends affect both groups similarly, a comparison group well matched on observable characteristics will produce unbiased estimates of program effects. This approach requires that differences on observable variables capture differences on unobserved variables as well. Regressions were estimated separately for Medicaid and Medicare beneficiaries, and pooled estimates were obtained by taking the weighted average of the Medicaid and Medicare estimates, where the weights reflect the relative size of the Medicaid/Medicare beneficiaries in the sample. Appendix A discusses the modeling approach and sample selection.

The evaluation defined the pre-enrollment period as the year before each participant enrolled in the program and the post-enrollment period as the following three years (though individuals who enrolled late in the period had a shorter follow-up period). The evaluation assigned a pseudo-enrollment date to each comparison beneficiary based on the enrollment date of the treatment group member to whom the comparison beneficiary was matched. The primary outcomes were total spending for Medicare beneficiaries and number of hospital admissions and number of ED visits for both Medicare and Medicaid beneficiaries. Total spending for Medicaid participants was not available because most Medicaid beneficiaries were enrolled in Medicaid managed care. Secondary outcomes included the number of primary care and specialty care visits. Appendix C provides full detailed results for the program on each outcome.

IMPACT RESULTS

The estimates do not suggest that the Rango integrated mobile platform and mobile application reduced service use among either Medicaid or Medicare beneficiaries (Table 4). According to the theory of action, participants' use of electronic self-care tools would improve their adherence to HIV treatment and their engagement in and satisfaction with care. This would ultimately reduce the costs associated with treatment failure and eliminate the need for more burdensome and expensive therapies, including hospitalizations and ED visits. However, for Medicaid beneficiaries, there was a slightly (7 percent) greater estimated *increase* for the treatment group than the comparison group in ED visits, although the estimate was not statistically significant. For Medicare beneficiaries, the pre-post changes in both hospitalizations and total Medicare spending were lower for the treatment group than for the comparison group (by 10 and 8 percent, respectively), but the sample is small and neither difference was statistically significant. Expenditure data were not available for the larger subgroup of Medicaid beneficiaries because most of them were in comprehensive managed care plans; however, the program did not generate reductions on hospitalizations or ED visits for Medicaid beneficiaries, so it is unlikely that the program reduced the plans' expenditures for these patients.

Sensitivity tests support the overall conclusion of no discernable program effects—there were no notable reductions in expenditures or service use when the analysis trimmed outliers or used a longer baseline period. Appendix C presents the full results of the impact analysis. Appendix D presents results from a Bayesian analysis.

Table 4. Estimated impact of Rango on selected outcome measures

	Medicare and Medicaid			Medicaid			Medicare		
	Impact	Percentage impact	*p*-value	Impact	Percentage impact	*p*-value	Impact	Percentage impact	*p*-value
Total Medicare expenditures ($ PBPM)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-166	-8.2%	0.37
Hospital stays, per 1,000 beneficiaries	-10.8	-2.1%	0.88	-2.3	< 1%	0.93	-52	-9.6%	0.47
ED and observation, visits per 1,000 beneficiaries	76.0	6.1%	0.43	88	6.8%	0.12	18	2.0%	0.85
Percentage of beneficiaries with a 30-day readmission	-0.7	-7.0%	0.71	-0.51	-4.9%	0.52	-1.9	-16.0%	0.32
Primary care visits	-99	-3.4%	0.61	-167	-6.4%	0.11	229	6.0%	0.42
Specialty care visits	218	2.1%	0.59	341	3.7%	0.29	-377	-2.3%	0.62
Sample sizes	2,952	10,819		2,532	8,873		420	1,946	

Sources: Mathematica's analysis of information from the awardee's program encounter database, and Medicare and Medicaid claims, encounter, and enrollment data from September 2012 through May 2018.

Notes: Impact estimates are based on the regression-adjusted difference between the randomized treatment and comparison group members. Percentage impacts are then calculated as the impact estimate divided by what the treatment group mean would have been absent the intervention (the treatment group mean in the post period minus the impact estimate). Appendix C presents the full impact estimates. Appendix D presents full impact estimates.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test;

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

n.a. = not available because most Medicaid beneficiaries were in comprehensive managed care plans;

The lack of favorable findings is not surprising given that the awardee recruited many beneficiaries who already adhered to their HIV medication regimens at enrollment. According to the program's theory of action, beneficiaries' use of electronic self-care tools should improve adherence to HIV treatment regimens, which in turn would reduce costs associated with treatment failures. The awardee's data indicated that beneficiary adherence improved during the program, and the Medicaid and Medicare claims data suggest that the treatment group was slightly (2 percent) more likely than the comparison group to fill a prescription during the follow-up period (not shown); however, the intervention had little opportunity to improve the outcomes for the 82 percent of beneficiaries who already adhered to HIV treatment at baseline.

VillageCare worked continuously to improve Rango, making important changes in the second year. Key changes included releasing the mobile application and adding new features that made the program more appealing to participants and easier to use. It is possible that these changes might have led to larger improvements in outcomes among participants who enrolled later in the program, but the results of the study did not differ among participants who enrolled early in the program versus those who enrolled late (see Appendix C, Table C.2).

CONCLUSION

Rango succeeded at engaging participants, with nearly 90 percent of participants using the mobile platform. However, there is no evidence that the program reduced expenditures or service use among either Medicaid or Medicare beneficiaries. The lack of a favorable impact is not surprising given that the goal of the Rango intervention was to improve treatment adherence and self-care among HIV-positive adults, and most participants were already adherent when they enrolled and managing their HIV disease. Thus, Rango could affect only the relatively small proportion of participants who were nonadherent at enrollment, who were not already effectively managing their HIV disease, and who were actively engaged in the program. To test whether an integrated mobile platform and mobile application for patients with HIV could improve treatment adherence and lower the use and cost of unnecessary services, recruitment would have to focus on beneficiaries who were not treatment adherent and who had difficulty managing their own health care.

Limitations of evaluation

The analysis has several limitations. First, the intervention's effect on treatment adherence will likely take years to generate observable impacts on the core service use and expenditure outcomes. Although Rango's program leaders and partners thought the evaluation could observe behavioral change (such as treatment adherence) in the short evaluation time frame, it would probably take more than a year or two for any behavioral changes to manifest as changes in health outcomes and health care costs. The average 26-month follow-up period used in this study was likely too short to observe such longer-term effects. Second, the evaluation calculated program impacts over participants only. If participants differed from eligible nonparticipants in ways that Medicare or Medicaid administrative files and claims could not capture, the results of

the study cannot be generalized to the full target population. Participants who already adhered to treatment were most likely to enroll and engage in Rango. This meant the intervention had little opportunity to improve the health care that participants received.

PROGRAM SUSTAINABILITY

VillageCare terminated the program when its award ended in February 2018 because it did not have other funding sources to sustain the program. However, the awardee continued to negotiate with Medicaid managed care organizations after the award ended in an effort to restart the program. When terminating the program, the awardee worked with program partners to identify alternative sources of services for program participants. The awardee also worked with its technology vendors to disable Rango's technology components, while maintaining functionality and data in the event of a future program reopening.

VillageCare's proposed payment model

VillageCare proposed to fund Rango services through a PBPM fee from health care entities that were responsible for the health of a patient population (such as accountable or managed care organizations). The awardee intended to allow payers to choose one of two levels of payment. The first included just the Rango software. The second, higher level of payment covered the software plus VillageCare staff who would run the program, such as the enrollment staff and health coaches.

The awardee's payment model featured a per beneficiary per month (PBPM) fee for each participant enrolled in the program. Payers would have been able to choose one of two levels of payment. The first included just the Rango software. The second, higher level covered the software plus VillageCare staff who would run the program, such as the enrollment staff and health coaches. The awardee waited to develop more specific details about its payment model, such as the payment amount, until it could demonstrate cost savings. The awardee assumed an entity responsible for the health of a patient population, such as an accountable care organization or a managed care organization, would make the payment.

This page has been left blank for double-sided copying.

Appendix A

Description of modeling strategy and analytic sample

This page has been left blank for double-sided copying.

The impact estimates for expenditures and number of visits or stays rely on a difference-in-differences approach with beneficiary fixed effects. They show the regression-adjusted change for the treatment group relative to that for the comparison group between baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or emergency department (ED) visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and whether the beneficiary had any hospital stay and any ED visit during the baseline period. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment (or pseudo-enrollment date for matched comparison beneficiaries). Regressions were estimated separately for Medicaid and Medicare beneficiaries, and pooled estimates were obtained by taking the weighted average of the Medicaid and Medicare estimates, where the weights reflect the relative size of the Medicaid/Medicare beneficiaries in the sample. Appendix A of Volume I of this report provides details on the general difference-in-differences modeling strategy and the standard set of outcomes.

The awardee defined participants as the 4,367 beneficiaries that enrolled in the Rango program from September 2012 through February 2017, as reported in the awardee's final encounter database. The impact analysis did not include all participants. Of the 4,367 participants, 211 could not be linked to the Medicaid or Medicare data, 955 were Medicare beneficiaries (including dual eligibles) and 3,201 were Medicaid beneficiaries. Of the 955 Medicare beneficiaries, 428 were from the study because they were enrolled in Medicare Advantage, 100 were excluded because of not having sufficient Medicare FFS data and 7 were excluded because they did not meet program eligibility criteria or were missing a key matching variable. Of the 3,201 Medicaid participants, 468 were excluded because they did not meet program eligibility criteria and 201 were excluded because they did not have complete Medicaid data. After making these exclusions, there were 420 Medicare and 2,532 Medicaid beneficiaries in the analysis.

Table A.1. Identification of final sample for impact analysis for VCC

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Number participants that could not be linked to Medicaid or Medicare data		**211**
Total Medicare program participants through August 31, 2017		**955**
Number dropped because not 18 years old (or older) or not a resident of New York City on day of enrollment	1	954
Number dropped because not alive at enrollment	0	954
Number dropped due to lack of Medicare FFS enrollment (Part A and B) on HCIA program enrollment date	87	867
Number dropped due to Medicare Advantage enrollment	428	439
Number dropped because Medicare was not primary payer on day of enrollment	0	439
Number dropped because did not have at least 90 days of Medicare FFS enrollment (Part A and B) in the baseline period	13	426

Table A.1 (*continued*)

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Number dropped because did not have HIV/AIDs diagnosis on or before day of enrollment	3	423
Number dropped because not alive 30 days after enrollment date	0	423
Number dropped in matching	3	420
Final analytic sample		**420**
Total Medicaid program participants through August 31, 2017		**3,201**
Number dropped because did not have HIV diagnosis on or before day of enrollment	305	2,896
Number dropped because not 18 or older or not a resident of New York City	163	2,773
Beneficiaries who died, had private insurance, restricted benefits, or were dually eligible for Medicare and Medicaid in the month of enrollment	173	2,560
Beneficiaries who lacked 90 days of Medicaid enrollment during baseline period	28	2,532
Final analytic sample		**2,532**

Sources: Mathematica's analysis of information from the awardee's program encounter database, and Medicare and Medicaid claims, encounter, and enrollment data from September 2012 through May 2018.

FFS = fee-for-service; HCIA = Health Care Improvement Awards; VCC = The Village Center for Care.

Appendix B

Results from balance assessment of treatment and comparison groups

This page has been left blank for double-sided copying.

Tables B.1 and B.2 show the variables used for matching the Medicare and Medicaid samples, respectively. Table B.1 displays the weighted means of baseline characteristics for the 420 treatment beneficiaries and the 1,946 matched comparison beneficiaries used in the Medicare impact analysis; Table B.2 displays the weighted means of baseline characteristics for the 2,532 treatment beneficiaries and the 8,873 matched comparison beneficiaries used in the Medicaid impact analysis. These tables show the means, difference in means, the percentage difference, and the standardized difference for each variable, which the evaluation calculated as the ratio of the difference in weighted means and the standard deviation of the variable (estimated on the treatment group). Standardized differences of less than 10 percent were generally considered a good fit.

The matching variables for the Medicare population included include demographic characteristics (age, gender, and race); Medicare entitlement and dual eligibility status; health status (as measured by the hierarchical condition category [HCC]) score and chronic condition indicators; Medicare expenditures in total and by type of service; and service use. The matching variables for the Medicaid population were similar to those used in the Medicare population, except Chronic Illness and Disability Payment System (CDPS) scores replaced HCC scores, and it used whether a beneficiary had a Medicaid claim for an HIV prescription at baseline as a matching variable. The analysis required an exact match on the quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries, and, for Medicaid beneficiaries, if they were enrolled in comprehensive managed care. The evaluation measured variables over various specified intervals within the 12 months before enrollment in the intervention. For more detail on the propensity score matching methodology used to identify the comparison group, see Appendix B in Volume I of this report.

The tables also show the results of the equivalency-of-means tests. *p*-values come from a weighted two-sample t-test, which provides evidence of the statistical significance of the difference in the means. The equivalence test *p*-values are the greater of two one-sided weighted t-test *p*-values equivalence tests, which assess whether the comparison group mean for a variable is more than 0.25 standard deviations away from the treatment group mean. The evaluation also performed an omnibus test in which the null hypothesis was that the treatment and matched comparison groups were balanced across all linear combinations of the covariates. It used the results to assess the closeness of fit between the treatment and matched comparison groups on key characteristics likely to be associated with study outcomes.

All of the differences between the treatment and matched comparison groups were small and statistically insignificant, except for the geographic distribution of Medicare sample members across counties.

Table B.1. Baseline characteristics of treatment and matched comparison groups for VillageCare: Medicare sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	54 (0.48)	54 (0.26)	-0.13 (0.77)	< +/-1	-0.01	0.87	< 0.01
Female, %	33 (2.3)	33 (1.1)	0.09 (3.3)	< +/-1	0.00	0.98	< 0.01
White, %	22 (2.0)	22 (0.96)	-0.32 (2.9)	-1.5	-0.01	0.91	< 0.01
Black, %	57 (2.4)	57 (1.1)	0.00 (3.4)	< +/-1	0.00	1.00	< 0.01
Hispanic, %	17 (1.8)	17 (0.85)	0.15 (2.6)	< +/-1	0.00	0.95	< 0.01
County of residence, %							
Bronx	31 (2.3)	31 (1.0)	0.10 (3.2)	< +/-1	0.00	0.98	< 0.01
Kings	20 (2.0)	20 (0.92)	0.00 (2.7)	< +/-1	0.00	1.00	< 0.01
Manhattan	37 (2.4)	37 (1.1)	0.00 (3.2)	< +/-1	0.00	1.00	< 0.01
Queens	9.8 (1.4)	9.8 (0.65)	0.00 (2.1)	< +/-1	0.00	1.00	< 0.01
Nassau	0.48 (0.34)	0.33 (0.14)	0.14 (0.42)	30	0.02	0.73	< 0.01
Richmond	1.4 (0.58)	1.4 (0.27)	0.05 (0.78)	3.3	0.00	0.95	< 0.01
Suffolk	0.24 (0.24)	0.14 (0.09)	0.10 (0.28)	40	0.02	0.74	< 0.01
Westchester	0.24 (0.24)	0.62 (0.18)	-0.38 (0.44)	-160	-0.08	0.39	0.03
Medicare entitlement and dual eligibility status, %							
Dually eligible for Medicare and Medicaid	92 (1.3)	93 (0.60)	-0.75 (1.8)	< +/-1	-0.03	0.68	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Original reason for Medicare entitlement: age	6.9 (1.2)	8.2 (0.64)	-1.3 (1.8)	-19	-0.05	0.48	< 0.01
Original reason for Medicare entitlement: disability	90 (1.5)	88 (0.75)	1.6 (2.2)	1.8	0.05	0.48	< 0.01
Part D coverage in the month before enrollment	97 (0.85)	97 (0.37)	-0.30 (1.1)	< +/-1	-0.02	0.79	< 0.01
Health status and diagnosis							
HCC score[a]	1.67 (0.05)	1.67 (0.03)	0.00 (0.07)	< +/-1	0.00	0.98	< 0.01
HIV-related prescription in the first 6 months of baseline, %	93 (1.2)	94 (0.57)	-0.50 (1.7)	< +/-1	-0.02	0.76	< 0.01
HIV-related prescription in the last 6 months of baseline, %	95 (1.1)	95 (0.50)	-0.54 (1.5)	< +/-1	-0.02	0.73	< 0.01
Medicare expenditures							
Total expenditures, year before enrollment (annualized)	1,394 (115)	1,348 (59)	47 (168)	3.4	0.02	0.78	< 0.01
Total expenditures, 3 months before enrollment (annualized)	1,527 (166)	1,611 (97)	-84 (271)	-5.5	-0.02	0.75	< 0.01
Service use							
0 hospitalizations, year before enrollment, %	79 (2.0)	79 (0.91)	-0.42 (2.9)	< +/-1	-0.01	0.89	< 0.01
1 hospitalization, year before enrollment, %	13 (1.7)	14 (0.77)	-0.31 (2.4)	-2.4	-0.01	0.90	< 0.01
2 hospitalizations, year before enrollment, %	4.3 (0.99)	4.4 (0.46)	-0.15 (1.4)	-3.5	-0.01	0.91	< 0.01
3 or more hospitalizations, year before enrollment, %	3.6 (0.91)	2.7 (0.37)	0.88 (1.2)	25	0.05	0.46	< 0.01
0 hospitalizations, 3 months before enrollment, %	91 (1.4)	91 (0.64)	0.38 (2.0)	< +/-1	0.01	0.85	< 0.01
1 hospitalization, 3 months before enrollment, %	7.1 (1.3)	7.4 (0.57)	-0.30 (1.8)	-4.2	-0.01	0.87	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
2 hospitalizations, 3 months before enrollment, %	1.7 (0.63)	1.7 (0.28)	-0.07 (0.90)	-4.0	-0.01	0.94	< 0.01
3 or more hospitalizations, 3 months before enrollment, %	0.24 (0.24)	0.25 (0.11)	-0.01 (0.34)	-5.0	0.00	0.97	< 0.01
Total hospitalizations for mental health or substance abuse (annualized; per 1,000 beneficiaries)	2.4 (0.74)	2.0 (0.32)	0.41 (1.0)	17	0.03	0.69	< 0.01
Total ED or observation visits[b] (annualized; per 1,000 beneficiaries)	831 (61)	821 (28)	10 (84)	1.2 Ha	0.01	0.90	< 0.01
Had ED or observation visit, 30 days before enrollment (percent)	6.7 (1.2)	6.8 (0.56)	-0.12 (1.7)	-1.7	0.00	0.95	< 0.01
Had ED or observation visit, 31 to 90 days before enrollment (percentage)	12 (1.6)	12 (0.71)	0.56 (2.2)	4.6	0.02	0.80	< 0.01
Had ED or observation visit with primary diagnosis for mental health or substance abuse (percentage)	3.1 (0.85)	2.8 (0.37)	0.25 (1.2)	8.1	0.01	0.83	< 0.01
Primary care visits in any setting[b] (annualized; per 1,000 beneficiaries)	4,052 (237)	3,948 (109)	104 (318)	2.6	0.02	0.74	< 0.01
Primary care visits in any setting, 30 days before enrollment[b] (annualized; per 1,000 beneficiaries)	5,243 (458)	5,173 (195)	70 (634)	1.3	0.01	0.91	< 0.01
Primary care visits in any setting, 31 to 90 days before enrollment[b] (annualized; per 1,000 beneficiaries)	4,306 (329)	4,072 (152)	234 (462)	5.4	0.03	0.61	< 0.01
Specialist visits in any setting[b] (annualized; per 1,000 beneficiaries)	13,265 (620)	13,103 (289)	161 (907)	1.2	0.01	0.86	< 0.01
Specialist visits in any setting, 30 days before enrollment[b] (annualized; per 1,000 beneficiaries)	15,380 (927)	15,483 (404)	-103 (1,336)	< +/-1	-0.01	0.94	< 0.01
Specialist visits in any setting, 31 to 90 days before enrollment[b] (annualized; per 1,000 beneficiaries)	13,666 (782)	13,687 (360)	-21 (1,144)	< +/-1	0.00	0.99	< 0.01
Had hospitalization, last 30 days of baseline, %	3.6 (0.91)	3.7 (0.38)	-0.16 (1.3)	-4.6	-0.01	0.90	< 0.01
Had outpatient visit, last 30 days of baseline, %	67 (2.3)	67 (1.1)	-0.31 (3.2)	< +/-1	-0.01	0.92	< 0.01

Table B.1 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Had physician visit, last 30 days of baseline, %	68 (2.3)	67 (1.1)	0.45 (3.3)	< +/-1	0.01	0.89	< 0.01
Had multiple visits with same clinician during year before enrollment, %	90 (1.5)	90 (0.70)	-0.15 (2.1)	< +/-1	0.00	0.94	< 0.01
Had physician, outpatient, or inpatient visit, 30 days before enrollment, %	80 (2.0)	80 (0.93)	0.01 (2.9)	< +/-1	0.00	1.00	< 0.01
Propensity score	-4.71 (0.05)	-4.71 (0.02)	0.00 (0.07)	< +/-1	0.00	0.98	< 0.01
Number of beneficiaries	420	1,946					
Omnibus test				Chi-squared statistic 14.91	Degrees of freedom 46.00	*p*-value 1.00	

Sources: Medicare claims and enrollment data from September 2012 through February 2017.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicare. Exact matching variables include the quarter of enrollment for treatment beneficiaries, the pseudo-enrollment date for the comparison beneficiaries, and enrollment in comprehensive managed care.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

ED = emergency department; HCC = hierarchical condition category; SE = standard error.

Table B.2. Baseline characteristics of treatment and matched comparison groups for VillageCare: Medicaid sample

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Demographics							
Age, years	47 (0.20)	47 (0.12)	0.23 (0.30)	< +/-1	0.02	0.45	< 0.01
Male, %	63 (0.96)	62 (0.52)	0.06 (1.4)	< +/-1	0.00	0.96	< 0.01
Hispanic, %	34 (0.94)	34 (0.50)	-0.06 (1.3)	< +/-1	0.00	0.96	< 0.01
County of residence, %							
Bronx	45 (0.99)	45 (0.52)	-0.04 (1.3)	< +/-1	0.00	0.97	< 0.01
Kings	24 (0.85)	24 (0.47)	0.12 (1.2)	< +/-1	0.00	0.92	< 0.01
New York	22 (0.82)	21 (0.42)	0.27 (1.2)	1.3	0.01	0.81	< 0.01
Queens	8.1 (0.54)	8.5 (0.31)	-0.37 (0.77)	-4.6	-0.01	0.63	< 0.01
Richmond	1.6 (0.25)	1.6 (0.14)	0.02 (0.35)	1.5	0.00	0.95	< 0.01
Medicaid enrollment, %							
Medicaid enrolled for 365 days	91 (0.57)	90 (0.25)	1.3 (0.82)	1.4	0.04	0.12	< 0.01
Enrolled in Medicaid fewer than 365 days of data in year before baseline	18 (0.76)	18 (0.33)	0.00 (1.1)	< +/-1	0.00	1.00	< 0.01
Health status and diagnosis							
CDPS score[a]	4.7 (0.04)	4.6 (0.02)	0.05 (0.05)	1.1	0.03	0.33	< 0.01
Any diagnosis of AIDS or HIV in the year before enrollment, %	98 (0.26)	98 (0.13)	-0.08 (0.37)	< +/-1	-0.01	0.84	< 0.01
Disabled, %	44 (0.99)	43 (0.52)	0.82 (1.5)	1.9	0.02	0.57	< 0.01
Pregnancy, %	1.1 (0.20)	1.1 (0.11)	-0.06 (0.29)	-5.2	-0.01	0.85	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Psychiatric condition, %	50 (0.99)	48 (0.52)	1.2 (1.4)	2.5	0.02	0.39	< 0.01
Substance abuse, %	37 (0.96)	37 (0.50)	-0.12 (1.3)	< +/-1	0.00	0.93	< 0.01
Service use							
0 hospitalizations, year before enrollment, %	76 (0.84)	76 (0.44)	0.53 (1.2)	< +/-1	0.01	0.66	< 0.01
1 hospitalization, year before enrollment, %	14 (0.70)	15 (0.36)	-0.75 (1.00)	-5.2	-0.02	0.45	< 0.01
2 hospitalizations, year before enrollment, %	4.8 (0.42)	4.6 (0.22)	0.14 (0.60)	3.0	0.01	0.81	< 0.01
3 or more hospitalizations, year before enrollment, %	4.6 (0.42)	4.5 (0.21)	0.08 (0.60)	1.7	0.00	0.89	< 0.01
0 hospitalizations, 3 months before enrollment, %	91 (0.56)	91 (0.29)	-0.02 (0.79)	< +/-1	0.00	0.98	< 0.01
1 hospitalization, 3 months before enrollment, %	6.8 (0.50)	6.8 (0.26)	0.04 (0.70)	< +/-1	0.00	0.96	< 0.01
2 hospitalizations, 3 months before enrollment, %	1.3 (0.23)	1.2 (0.12)	0.10 (0.32)	7.4	0.01	0.76	< 0.01
3 or more hospitalizations, 3 months before enrollment, %	0.51 (0.14)	0.63 (0.08)	-0.12 (0.22)	-23	-0.02	0.58	< 0.01
Had hospitalization, 30 days before enrollment, %	2.4 (0.30)	2.2 (0.17)	0.21 (0.43)	8.8	0.01	0.62	< 0.01
Had hospitalization for mental health reasons, principal diagnosis, %	7.1 (0.51)	7.2 (0.26)	-0.13 (0.71)	-1.9	-0.01	0.85	< 0.01
Total ED or observation visits[b] (annualized; per beneficiary)	1.0 (0.03)	0.98 (0.02)	0.04 (0.04)	3.6	0.02	0.40	< 0.01
Had ED or observation visit, 30 days before enrollment, %	8.3 (0.55)	8.0 (0.28)	0.29 (0.74)	3.5	0.01	0.70	< 0.01
Had ED or observation visit, 31 to 90 days before enrollment, %	14 (0.70)	14 (0.35)	0.57 (0.94)	3.9	0.02	0.55	< 0.01
Had any ED or observation visit with mental health diagnosis as principal diagnosis, %	6.4 (0.49)	6.1 (0.25)	0.33 (0.68)	5.1	0.01	0.63	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Primary care visits, any setting[b] (per beneficiary)	2.8 (0.08)	2.8 (0.04)	-0.03 (0.12)	-1.0	-0.01	0.81	< 0.01
Primary care visits in any setting, 31 to 90 days before enrollment (per beneficiary)	0.44 (0.02)	0.44 (0.01)	0.00 (0.03)	< +/-1	0.00	0.92	< 0.01
Primary care visits in any setting, 30 days before enrollment[b] (per beneficiary)	0.24 (0.01)	0.25 (0.01)	-0.01 (0.02)	-2.4	-0.01	0.71	< 0.01
Specialist visits, any setting[b] (per beneficiary)	8.6 (0.19)	8.5 (0.10)	0.07 (0.27)	< +/-1	0.01	0.80	< 0.01
Specialist visits in any setting, 30 days before enrollment[b] (per beneficiary)	0.67 (0.02)	0.63 (0.01)	0.04 (0.03)	5.7	0.03	0.25	< 0.01
Specialist visits in any setting, 31 to 90 days before enrollment[b] (per beneficiary)	1.4 (0.04)	1.3 (0.02)	0.05 (0.06)	3.5	0.02	0.44	< 0.01
Had any 30-day unplanned readmission during year before enrollment, %	4.3 (0.40)	4.1 (0.20)	0.21 (0.56)	4.9	0.01	0.71	< 0.01
Had physician visit 30 days before enrollment, %	45 (0.99)	44 (0.52)	0.63 (1.4)	1.4	0.01	0.66	< 0.01
Had two or more visits with the same provider during year before enrollment, %	75 (0.86)	75 (0.46)	0.04 (1.2)	< +/-1	0.00	0.98	< 0.01
Had one or more physician, ED, or hospital visit 30 days before enrollment, %	46 (0.99)	46 (0.53)	0.58 (1.4)	1.2	0.01	0.69	< 0.01
Pharmacy claims							
Any Rx claims for HIV drugs, 1 to 6 months before enrollment, %	93 (0.51)	93 (0.28)	0.28 (0.73)	< +/-1	0.01	0.71	< 0.01
Any Rx claims for HIV drugs, 7 to 12 months before enrollment, %	87 (0.66)	86 (0.34)	1.4 (0.97)	1.6	0.04	0.14	< 0.01
Any Rx claims for HIV drugs, 13 to 18 months before enrollment, %	83 (0.74)	81 (0.38)	1.7 (1.1)	2.0	0.04	0.12	< 0.01
Any Rx claims for HIV drugs, 19 to 24 months before enrollment, %	80 (0.79)	79 (0.41)	1.3 (1.1)	1.7	0.03	0.24	< 0.01
Number of Rx claims for HIV drugs, 13 to 18 months before enrollment	10 (0.16)	10.0 (0.09)	0.17 (0.23)	1.7	0.02	0.46	< 0.01
Number of Rx claims for HIV drugs, 19 to 24 months before enrollment	10 (0.16)	9.8 (0.09)	0.17 (0.25)	1.7	0.02	0.48	< 0.01

Table B.2 (*continued*)

Characteristic	Treatment mean (SE)	Matched comparison mean (SE)	Adjusted difference (SE)	Percentage difference	Standardized difference	t-test *p*-value	Equivalence *p*-value
Number of Rx claims for HIV drugs, 1 to 6 months before enrollment	10 (0.14)	10 (0.08)	0.07 (0.21)	< +/-1	0.01	0.72	< 0.01
Number of Rx claims for HIV drugs, 7 to 12 months before enrollment	10 (0.15)	9.9 (0.08)	0.24 (0.23)	2.4	0.03	0.29	< 0.01
Propensity score	0.13 (0.00)	0.13 (0.00)	0.00 (0.00)	< +/-1	0.00	0.86	< 0.01
Number of beneficiaries	2,532	8,876					
Omnibus test				Chi-squared statistic	Degrees of freedom	P-value	
				612.60	118.00	0.00	

Sources: Medicaid claims and encounter data between September 2012 and February 2017. Medicaid data used for these analyses were interim T-MSIS analytic files; these files were made available in the Chronic Conditions Warehouse for the purposes of this evaluation. Because the data were not final, findings might not be replicable with the newly available TAF research identifiable files or other data sources.

Note: Standard errors in parentheses. Standardized difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a weighted two-sample t-test; equivalence test *p*-values are the greater of the *p*-values for the two one-sided weighted t-tests of whether the true treatment–comparison difference exceeded 0.25 standard deviations of the variable. The comparison group means in the table are calculated by weighting observations by the matching weight. The matching weight reflects the number of times a comparison beneficiary is matched to a treatment beneficiary. Unlike the weight used in the baseline characteristics table in the body of the report and the model results tables in the body of the report and Appendix C, the matching weight does not account for the number of months a beneficiary was enrolled in Medicaid. Exact matching variables include the quarter of enrollment for treatment beneficiaries and pseudo-enrollment date for the comparison beneficiaries.

[a] The CDPS score is based on demographic characteristics and the presence of specific diseases and conditions characterized by expected cost (ranging from extra high to very low). A score above 1 indicates higher-than-average expected spending, and a score below 1 indicates lower-than-average spending

[b] Top-coded at the 98th percentile based on the distribution of the treatment beneficiaries in the baseline and follow-up periods.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; FFS = fee-for-service; HCBS = home and community-based services; HCC = hierarchical condition category; SE = standard error.

This page has been left blank for double-sided copying.

Appendix C

Detailed results from impact estimates

This page has been left blank for double-sided copying.

Tables C.1 and C.2 display the results from the impact analysis. Table C.1 shows the impact estimates for the combined sample of Medicaid and Medicare beneficiaries. Table C.2 shows the impact estimates for Medicaid and Medicare beneficiaries, measured separately over intervention Years 1 through 3. The evaluation estimated the models over Medicare expenditures, number of services used (per 1,000 beneficiaries), and probability of using any service, in total and by type of service. The estimated percentage impact of the program is the estimated impact divided by a counterfactual value defined as the treatment group mean minus the impact estimate. One, two, or three asterisks indicate impact estimates that differ statistically from zero at the .10, .05, and .01 levels, respectively, using a two-tailed test.

Table C.1. Estimated impact of the VillageCare intervention on selected use measures during one-, two-, and three-year follow-up periods

	Medicare and Medicaid, combined sample		
	Impact estimate (SE)	Percentage impact[a]	*p*-value
Hospital stays, per 1,000 beneficiaries			
Baseline year			
Year 1	-5.7 (41)	-1.1%	0.89
Year 2	-18.4 (46)	-3.3%	0.69
Year 3	14.8 (102)	2.5%	0.88
Cumulative	-10.8 (38)	-2.1%	0.78
ED or observation visits, per 1,000 beneficiaries			
Baseline year			
Year 1	75.0 (70)	6.1%	0.28
Year 2	63.8 (86)	5.1%	0.46
Year 3	83.5 (138)	6.2%	0.54
Cumulative	76.0 (66)	6.1%	0.25
Percentage of beneficiaries with any hospital admission in a time period			
Baseline year			
Year 1	0.3 (1.0)	1.4%	0.80
Year 2	-0.2 (1.0)	-0.7%	0.91
Year 3	-0.6 (3.0)	-2.8%	0.80
Cumulative	1.6 (2.0)	3.9%	0.30
Percentage of beneficiaries with any ED or observation visits in a time period			
Baseline year			
Year 1	5.7** (1.0)	13%	0.00
Year 2	4.4 (2.0)	10%	0.01
Year 3	1.4 (3.0)	3.5%	0.61
Cumulative	5.0** (1.0)	7.2%	0.00
Primary care visits in ambulatory setting, per 1,000 beneficiaries			
Baseline year			
Year 1	31.1 (154)	1.1%	0.84

Table C.1 (*continued*)

	Medicare and Medicaid, combined sample		
	Impact estimate (SE)	Percentage impact[a]	*p*-value
Year 2	-206.7 (195)	-7.1%	0.29
Year 3	-596.2** (265)	-23%	0.02
Cumulative	-99 (151)	-3.4%	0.51
Specialist visits in all settings, per 1,000 beneficiaries			
Baseline year			
Year 1	61.1 (425)	0.6%	0.89
Year 2	338.4 (560)	3.0%	0.55
Year 3	307.0 (1032)	2.6%	0.77
Cumulative	217.6 (431)	2.1%	0.61
Percentage of beneficiaries with a hospital readmission			
Baseline year			
Year 1	0.0 (1.0)	0.5%	0.97
Year 2	-0.4 (1.0)	-8.4%	0.57
Year 3	-0.7 (1.0)	-15%	0.60
Cumulative	-0.7 (1.0)	-7.0%	0.48
Sample sizes			
Number of beneficiaries			
Year 1	13,771		
Year 2	11.361		
Year 3	5,009		
Cumulative	13,771		

Sources: Mathematica's analysis of information from the awardee's program encounter database, and Medicare and Medicaid claims, encounter, and enrollment data from September 2012 through May 2018.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FFS = fee-for-service.

Table C.2. Estimated impact of the VillageCare intervention on selected Medicare FFS expenditures (dollars PBPM) and Medicare and Medicaid use measures during one-, two-, and three-year follow-up periods

	Medicare					Medicaid				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Total expenditures ($ PBPM)										
Baseline year	1,371	1,341								
Year 1	1,884	1,856	-2.1 (211)	< 1%	0.99					
Year 2	1,852	2,223	-401* (210)	-19%	0.06					
Year 3	2,141	2,419	-308 (484)	-14%	0.52					
Cumulative	1,872	2,008	-166 (187)	-8.2%	0.37					
Hospital stays, per 1,000 beneficiaries										
Baseline year	400	343				471	478			
Year 1	500	458	-16 (73)	-3.1%	0.83	500	511	-3.6 (30)	0%	0.9
Year 2	535	584	-107 (84)	-19%	0.2	538	545	-0.66 (34)	0%	0.98
Year 3	629	566	5.7 (197)	1.0%	0.98	591	581	17 (59)	3.1%	0.77
Cumulative	507	502	-52 (71)	-9.6%	0.47	516	525	-2.3 (27)	0%	0.93
ED or observation visits, per 1,000 beneficiaries										
Baseline year	920	849				1,162	1,154			
Year 1	1,067	863	133 (-102)	15%	0.19	1,386	1,314	63 (-61)	4.8%	0.3
Year 2	878	928	-122 (-131)	-12%	0.35	1,445	1,335	101 (-74)	7.8%	0.17
Year 3	930	1,002	-143 (-148)	-15%	0.33	1,610	1,462	139 (-135)	11%	0.3
Cumulative	986	896	18 (98)	2.0%	0.85	1,436	1,339	88 (57)	6.8%	0.12
Percentage of beneficiaries with any hospital admission in a time period										
Baseline year	22	21				24	25			
Year 1	22	23	-1.4 (-2.2)	-5.9%	0.54	24	24	0.68 (-0.98)	2.9%	0.49
Year 2	22	26	-3.8 (-2.6)	-15%	0.15	25	24	0.57 (-1.1)	2.4%	0.61
Year 3	21	25	-4.5 (-4.3)	-18%	0.3	23	22	0.31 (-1.9)	1.4%	0.87
Cumulative	45	46	-0.54 (-2.8)	-1.2%	0.85	44	42	2.1* (-1.2)	5.0%	0.08

Table C.2 (*continued*)

	Medicare					Medicaid				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Percentage of beneficiaries with any ED or observation visits in a time period										
Baseline year	44	42				47	43			
Year 1	52	39	13.0*** (-2.6)	32%	0.001	50	46	4.2*** (-1.1)	9.3%	0.001
Year 2	43	43	-0.4 (-3)	0%	0.89	51	46	5.4*** (-1.3)	12%	0.001
Year 3	46	44	1.5 (-4.7)	3.3%	0.76	41	40	1.4 (-2.1)	3.4%	0.52
Cumulative	75	67	7.8*** (-2.2)	12%	0.001	74	69	4.4*** (-0.99)	6.3%	0.001
Primary care visits in ambulatory setting, per 1,000 beneficiaries										
Baseline year	3,524	3,570				2,412	2,348			
Year 1	4,081	3,898	228 (-281)	5.9%	0.42	2,667	2,612	-9.7 (-110)	0%	0.93
Year 2	4,266	4,049	264 (-373)	6.9%	0.48	2,186	2,423	-301** (-134)	-12%	0.03
Year 3	3,650	3,938	-242 (-458)	-6.0%	0.6	1,444	2,062	-683*** (-189)	-28%	0.001
Cumulative	4,129	3,946	229 (-283)	6.0%	0.42	2,419	2,521	-167 (-104)	-6.4%	0.11
Specialist visits in all settings, per 1,000 beneficiaries										
Baseline year	13,515	13,315				9,141	9,272			
Year 1	15,146	15,430	-483 (-768)	-3.1%	0.53	9,007	8,965	174 (-309)	2.0%	0.57
Year 2	16,001	16,084	-283 (-1004)	-1.7%	0.78	10,303	9,972	463 (-417)	4.8%	0.27
Year 3	17,536	17,429	-93 (-1815)	0%	0.96	10,422	10,148	405 (-721)	3.7%	0.57
Cumulative	15,552	15,730	-377 (-764)	-2.3%	0.62	9,542	9,333	341 (-321)	3.7%	0.29
Percentage of beneficiaries with a hospital readmission										
Baseline year	3.0	3.0				4.0	4.0			
Year 1	5.0	4.0	0.66 (-1.2)	16%	0.57	4.0	4.0	-0.11 (-0.46)	-2.6%	0.81
Year 2	5.0	6.0	-1.7 (-1.4)	-27%	0.22	5.0	5.0	-0.19 (-0.59)	-3.8%	0.74
Year 3	4.0	5.0	-1.5 (-2.2)	-29%	0.5	4.0	4.0	-0.48 (-0.95)	-11%	0.61
Cumulative	10	12	-1.9 (-1.9)	-16%	0.32	10	10	-0.51 (-0.78)	-4.9%	0.52

Table C.2 (*continued*)

	Medicare					Medicaid				
	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value	Treatment group mean	Comparison group mean	Impact estimate (SE)	Percentage impact[a]	*p*-value
Sample sizes										
Number of beneficiaries										
Year 1	420	1,946				2,532	8,873			
Year 2	326	1,571				2,140	7,324			
Year 3	168	818				913	3,110			
Cumulative	420	1,946				2,532	8,873			

Sources: Mathematica's analysis of information from the awardee's program encounter database, and Medicare and Medicaid claims, encounter, and enrollment data from September 2012 through May 2018.

Note: Impact estimates for expenditures and number of visits or stays are based on a difference-in-differences approach and show the regression-adjusted change for the treatment group relative to that for the comparison group between the baseline and intervention periods. The impact estimate for the binary outcomes of any hospital stay or ED visit is a regression-adjusted treatment–comparison difference based on a cross-sectional regression that controls for a beneficiary's characteristics and the probability of having any hospital stay or ED visit at baseline. The intervention years are beneficiary specific and defined relative to each beneficiary's date of enrollment.

[a] Percentage impact is relative to a counterfactual value defined as the treatment group mean minus the impact estimate.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

Appendix D

Results from Bayesian analysis

This page has been left blank for double-sided copying.

In addition to the traditional frequentist analysis presented in the body of this report, the program impacts for the Village Center for Care (VCC) were also estimated using a Bayesian approach. The Bayesian approach supplements the main analysis by framing conclusions in probabilistic terms, which facilitates decision making by summarizing both the size and the certainty of an impact in a single value. To draw probabilistic conclusions, external or prior evidence is required. In this analysis, the findings from the evaluation of 87 awardees included in the first round of the Health Care Innovation Awards (HCIA R1) provided the prior evidence, with more weight on results from awardees with background characteristics similar to VCC. Probabilities were calculated using the results of a Bayesian regression that jointly models impacts on CMS's four core outcomes, thereby improving the precision of the impact estimates. For more detail on the Bayesian methodology, see Appendix D in Volume I of this report.

Table D.1 compares the Bayesian impact estimates for CMS's four core outcomes with the regression estimates obtained from the frequentist analysis reported in the body of this report. Combining prior evidence from HCIA R1 with the estimates from the frequentist regressions for VCC led to a Bayesian estimate of the program's impact on total Medicare expenditures of -4 percent (an estimated reduction of $79 per beneficiary per month) across the first three program years.

Table D.1. Comparison of frequentist and Bayesian impact estimates for VCC in the first three years after enrollment

		Impact estimate (95 percent interval)			Percentage impacts	
Payer	Outcome	Frequentist	Bayesian	Prior	Frequentist	Bayesian
Medicaid	Hospital admissions	-2.3 (-56, 52)	-6.8 (-43, 28)	-2%	> -1%	-1%
	ED visits	88 (-24, 200)	-23 (-119, 68)	-2%	7%	-2%
	Readmissions	-0.51 (-2.0, 1.0)	-0.15 (-0.86, 0.55)	-2%	-5%	-1%
Medicare	Total expenditures ($ PBPM)	-166 (-532, 200)	-79 (-246, 84)	-2%	-8%	-4%
	Hospital admissions	-52 (-192, 88)	-20 (-67, 25)	-2%	-9%	-4%
	ED visits	18 (-174, 211)	-39 (-120, 39)	-2%	2%	-4%
	Readmissions	-1.9 (-5.6, 1.8)	-0.46 (-1.4, 0.54)	-2%	-16%	-4%
Pooled	Hospital admissions	-11 (-86, 64)	-8.8 (-42, 23)	-2%	-2%	-2%
	ED visits	76 (-53, 205)	-25 (-109, 55)	-2%	6%	-2%
	Readmissions	-0.7 (-1.5, 4.8)	-0.19 (-0.87, 0.43)	-2%	-7%	-2%

Source: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS and Medicaid claims from September 2012 to May 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Notes: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending and are evaluated in the Medicare sample only. Readmissions are the percentage of beneficiaries with a readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

Table D.1 (*continued*)

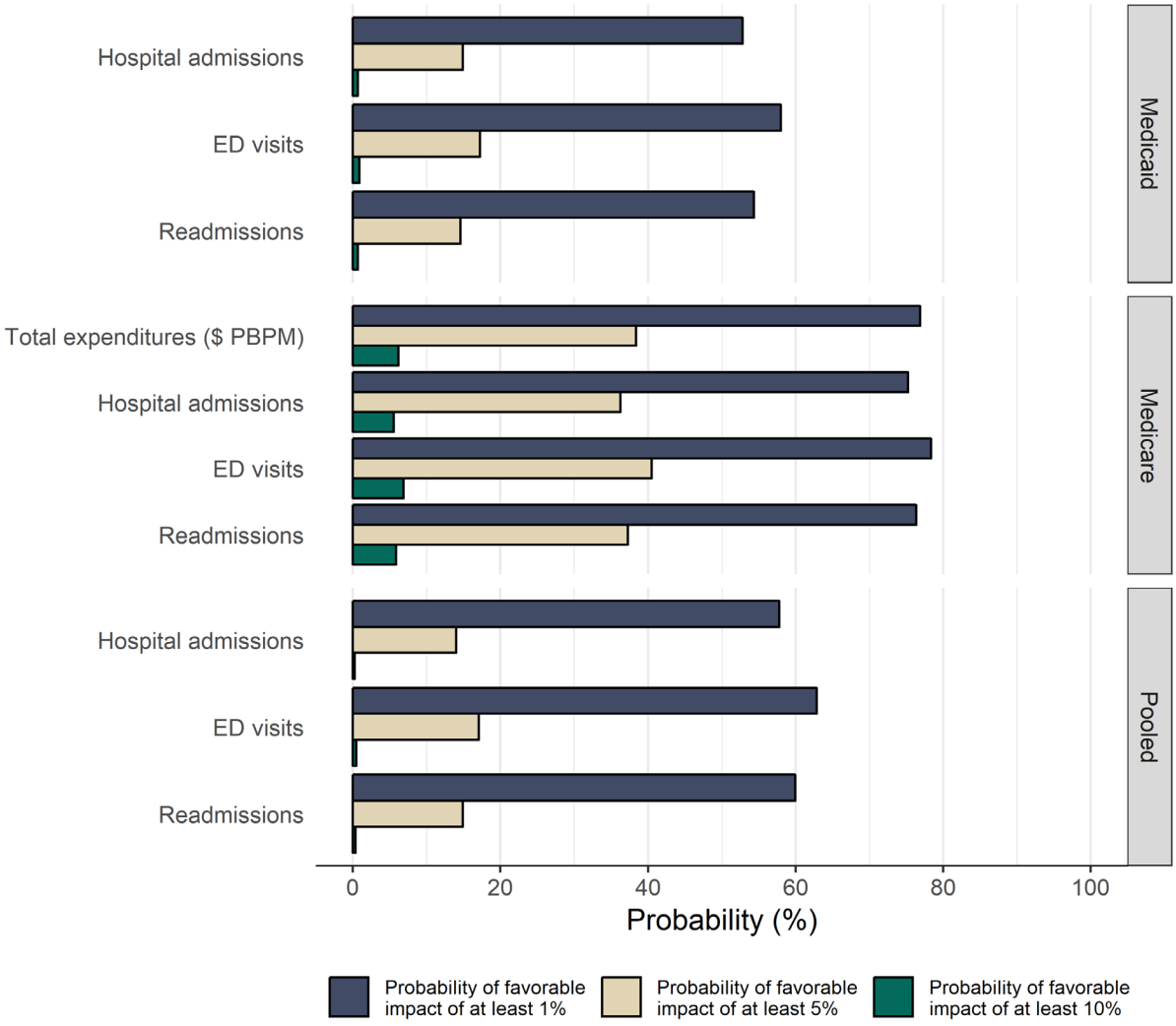
Intervals for frequentist analysis results are traditional confidence intervals, calculated using the standard error of the impact estimate. Bayesian intervals are credible intervals calculated as the 2.5 and 97.5 quantiles of the posterior distribution for the impact.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

Because the frequentist results are imprecise, the Bayesian model gave more weight to the prior and produced more neutral estimates. Despite these differences, the Bayesian results substantively agree with the frequentist results in finding that all impacts are statistically indistinguishable from zero.

To determine whether to continue the program, it is useful to know whether the estimated impacts correspond to high probabilities of achieving policy targets, such as a 5 percent reduction in expenditures. Figure D.1 shows the probability that VCC achieved favorable impacts across the first three program years on CMS's four core outcomes at three different thresholds: (1) a favorable impact of 1 percent or more, (2) a favorable impact of 5 percent or more, and (3) a favorable impact of 10 percent or more.

Figure D.1. Probability that the VCC program had a favorable impact on key outcomes



Source: Mathematica's analysis of awardee-provided enrollment data and Medicare FFS and Medicaid claims from September 2012 to May 2018. The Bayesian analysis also incorporated HCIA R1 meta-analysis data.

Note: ED visits include observation stays. Total expenditures include both Medicare Parts A and B spending and are evaluated in the Medicare sample only. Readmissions are the percentage of beneficiaries with a readmission. The Bayesian regression also incorporates assumptions about the likely distribution of impact estimates; these assumptions are based on data from the HCIA R1 evaluation.

ED = emergency department; HCIA = Health Care Innovation Awards; PBPM = per beneficiary per month.

There is a modest probability—in the range of 60 percent—that VCC had a favorable impact of 1 percent or more on hospital admissions, emergency department visits, and readmissions in the pooled sample. Probabilities are higher, at close to 80 percent, in the Medicare sample, suggesting that favorable effects are concentrated in this population. Though they are promising, these probabilities are not large enough to indicate a substantial impact. Thus, the Bayesian analysis corroborates the findings from the frequentist analysis that the VCC program did not have a meaningful impact on total expenditures or service utilization.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Washington University School of Medicine in St. Louis

September 2020

Carol Urato, Margaret Gerteis, Stefanie Pietras, Bob Schmitz, and Boyd Gilman

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

WASHINGTON UNIVERSITY SCHOOL OF MEDICINE IN ST. LOUIS

The Washington University School of Medicine received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to create the Contraceptive Choice Center (C3), an innovation designed to reduce unintended pregnancy among high-risk women in the St. Louis, Missouri, metropolitan area by improving access to effective methods of contraception. The C3 program, launched in January 2015, sought to engage reproductive-age women 14 years and older at risk for unintended pregnancy and childbirth. The HCIA R2-funded intervention period ended in August 2017. Washington University received a three-month no-cost extension through November 2017 that enabled it to continue to enroll new patients, track existing patients as part of its own research study, and pursue its payment model with payers. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- The C3 enrolled participants and delivered services consistent with the CHOICE model—reducing barriers to access, engaging patients through contraceptive counselling, and delivering clinical care with high rates of LARC uptake.
- A rigorous impact analysis of the C3 program's impacts on unintended pregnancy and associated Medicaid costs was not feasible because of the target population's prior exposure to the CHOICE program, lack of comparative data about pregnancy intention in Medicaid claims, and the inability to detect cost impacts related to childbirth within the period of the evaluation.
- This descriptive analysis, based on the awardee's self-measurement and monitoring data for 3,022 direct participants and on Title X public use data, should therefore not be interpreted as an evaluation of program impacts.

The C3 program was modelled after the Contraceptive CHOICE Project, a program supported by private funding that combined comprehensive contraceptive counseling and support with same-day access to long-acting reversible contraceptive (LARC) methods. Through HCIA R2, Washington University sought to implement the model using Medicaid, commercial insurance, and Title X family planning funding, and to develop a payment model that would reduce barriers to access by incentivizing providers to adopt the approach. The three major components of the C3 service delivery model included (1) training a social worker to serve as an insurance navigator (patient navigation); (2) using nonclinician health educators as contraceptive counselors (patient engagement); and (3) providing same-day contraceptive clinical services, including LARC insertion (direct care provision). The program aimed to increase uptake and continuation of the most effective contraceptive methods, reduce rates of unintended pregnancy, and reduce costs associated with unintended pregnancy and births.

A rigorous evaluation of C3's impacts on unintended pregnancy rates and associated costs was not possible for three reasons. First, prior experience with the CHOICE program in St. Louis

precluded identifying a baseline population lacking exposure to the intervention that could serve as a comparison group. Second, it was not possible to determine whether pregnancy outcomes in a comparison group were intended or unintended using Medicaid claims data. Third, downstream costs associated with unintended pregnancy and childbirth would not be discernible through claims data within the period of this evaluation. This descriptive analysis of participants' enrollment and engagement and program implementation relies on the awardee's self-measurement and monitoring reports for 3,022 direct participants and on Title X public use data. For these reasons, the descriptive results presented in this report should not be interpreted to imply that the program had a causal impact on outcomes

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	Washington University implemented the C3 to provide evidence-based contraceptive counseling and family planning services, including same-day insertion of LARCs (for example, intrauterine devices and implants) to reproductive-age women in St. Louis, Missouri.
Major innovation	The C3 was innovative because it entailed implementing the CHOICE model of comprehensive contraceptive care using Medicaid, commercial insurance, and Title X family planning funding in lieu of private funding. By demonstrating the cost- effectiveness of the model to payers, Washington University proposed to develop a payment model that would incentivize providers to adopt the approach.
Program components	• Patient navigation to help patients review insurance options and apply for coverage • Patient engagement to give structured, evidence-based contraceptive counseling and support to all patients • Direct care provision to provide same-day contraceptive services that follow evidence-based guidelines
Target population	The C3 sought to engage reproductive-age women 14 years and older in the St. Louis area who were at risk for unintended pregnancy and childbirth.
Total enrollment	Washington University enrolled 3,022 direct participants in the C3 program (75 percent of its revised enrollment projection) by the end of the three-year cooperative agreement. Direct participants included patients who met eligibility requirements and consented to receive clinical services during their first visit.
Theory of change or theory of action	Washington University hypothesized that reducing barriers to evidence-based methods of contraception among women of childbearing age and among clinicians who provide family planning services would increase the uptake of methods proven to be most effective, resulting in a reduction of unintended pregnancies and childbirth and their associated costs.
Award amount	$4,034,879
Effective launch date	The program began operating on January 8, 2015, five months after the award date.
Program settings	The C3 was established as a dedicated clinic housed in the Division of Clinical Research, Washington University Department of Obstetrics and Gynecology, staffed with social workers, nonclinician health educators, and clinicians trained to provide contraceptive services using evidence-based guidelines.
Market area	St. Louis, Missouri, metropolitan area
Target outcomes	• Increase in uptake of LARCs to 50 percent of new contraceptive methods • Increase in contraceptive continuation and satisfaction • Reduce rate of unintended pregnancy by 10 percent • Reduce costs associated with unintended births by 15 percent

Table 1 (*continued*)

Program characteristics	Description
Payment model	By the end of the award, Washington University had prepared detailed memoranda proposing a bundled episode-based payment model for contraceptive care to Missouri HealthNet and Medicaid managed care plans. However, the awardee had been unable to negotiate agreements with payers.
Sustainability plans	Washington University's long-term plan for sustaining the C3 model relied on Missouri HealthNet's approval of its proposed bundled payment model and Medicaid managed care plans' adoption of a similar payment model. The awardee also focused on generating revenue from other sources, through participation in Title X, expanding billable services for commercially insured patients, and through private fund raising. However, the awardee doubted C3 could sustain services for the uninsured without Title X funding or contraceptive counselling without reimbursement under a new payment model.

C3 = Contraceptive Choice Center; LARC = long-acting reversible contraceptive.

PROGRAM DESIGN AND ADAPTATION

Washington University designed the C3 clinic to deliver services in a manner consistent with the earlier CHOICE model. This included educating patients about contraceptive options, evidence-based contraceptive services (including same-day access to LARCs), and follow-up support. To address patients' barriers to care and adapt the model from a research setting to the real world, the awardee added insurance navigation and counseling as core components of its HCIA R2 C3 service delivery model.[1] Missouri's early decision not to expand its Medicaid program and (later) to withdraw from the federally funded Medicaid family planning waiver demonstration prompted the awardee to devote more attention to financial barriers to access than initially anticipated and to revise its initial enrollment targets downward.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Washington University's prior experience with CHOICE gave it the fundamental model of staffing, training, patient flow, and service delivery needed to effectively implement the C3 program. Staff reported little change, and few challenges, to the core work of engaging patients in contraceptive counseling and providing contraceptive care. The level of commitment to C3's mission was also evident in interviews with program staff. Participating clinicians (nurse practitioners and physicians) and other C3 staff were consistently positive in their assessments of the program's effects on the delivery of patient care. Almost all respondents cited buy-in from clinicians and staff and high levels of patient engagement as key factors associated with C3's perceived effects on the delivery of care.

However, the awardee faced numerous challenges attracting and enrolling participants and implementing the program to effectively address financial barriers to access. Several factors

[1] The Third Annual Evaluation report provides additional details on the design and implementation of the C3 program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

influenced Washington University's inability to meet its three-year enrollment goals. It took time and continuous effort to establish a reliable referral network among providers, but by the third year of the cooperative agreement such referrals accounted for about 15 to 20 percent of new clinic patients. Community organizations serving high-risk populations responded favorably to C3's informational outreach about contraceptive care, but did not actively refer clients to the clinic if pregnancy prevention was not an organizational priority. A lack of awareness of the program's services among the target population also hindered enrollment efforts, but word of mouth, digital media outreach, and radio advertising proved somewhat effective, over time, in attracting new patients to the clinic. Missouri's limits on Medicaid eligibility and changing state policy on women's reproductive health services further constrained C3's outreach efforts. The awardee subsequently revised its enrollment targets downward to be more in line with its actual enrollment experience.

Implications of program implementation for achieving program goals

- Prior experience with CHOICE provided the basic model of staffing, training, patient flow, and service delivery.
- Missouri's decision (1) not to expand Medicaid eligibility immediately before the program start, combined with its later decisions to (2) withdraw from the federal Medicaid family planning waiver demonstration and (3) exclude the awardee from the state's Uninsured Women's Health Services Program raised financial barriers to access that required changes to service delivery.
- The clinic's Title X status enabled the awardee to provide services at little or no cost to uninsured patients and to obtain 340b discounted drug pricing, which enabled them to keep LARCs in stock for same-day insertion.

In addition, the state's Medicaid policy and eligibility limits required the awardee to cobble together a viable support system to reduce women's financial barriers to access. C3's status as a Title X clinic was critical to its ability to provide services at reduced or no cost to patients who lacked insurance and gave the clinic access to LARCs at reduced price. Keeping up to speed with the complicated and changing insurance landscape and third-party billing requirements reportedly took its toll on C3 staff, who credited Washington University's billing infrastructure with helping to overcome this challenge. Awardee leadership also acknowledged having underestimated the number of staff needed to carry out basic enrollment, registration, prior approval, documentation, and reporting tasks.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

The descriptive analysis presented in this report presents outcomes on participants' enrollment and engagement and program implementation based on Washington University's self-measurement and monitoring reports (including data gathered for C3's research study), Title X public use data, and interviews with C3 staff. Washington University considered all women who

qualified and consented to participate in the C3 research study or received clinical services during their first visit to be direct participants. The quantitative results presented here relied on direct reports from participants.

Enrolling and engaging participants

Washington University was partially successful in achieving its enrollment targets. The awardee revised its initial target of 10,000 direct enrollees downward, after Missouri elected not to expand its Medicaid program, withdrew from the federally funded Medicaid family planning waiver demonstration, and chose not to cover Washington University under the state-funded Uninsured Women's Health Services Program that replaced it. The awardee reported enrolling 3,022 direct participants in its C3 program from January 2015 (when it launched its program) through August 2017, which represents about 75 percent of its 4,047 adjusted three-year projected goal.

In interviews, C3's nonclinician contraceptive counselors reported reviewing contraceptive options with patients (in English or Spanish) and answering their questions, following standard protocols, before eliciting their preferences. By design, counselors spent more time with adolescents and with women who appeared to have limited prior understanding of reproductive physiology or how different contraceptive methods worked. The clinic's monthly self-monitoring statistics indicated uniformly high levels of patient satisfaction with C3 counselors (10.5 points or higher in most months, out of 11.0 possible points).

Because many insurers (including Medicaid) required prior approval for LARC insertion, C3 staff also sought to secure such approval before scheduled appointments to avoid delays in treatment. According to the awardee's monthly self-monitoring statistics across all program years, C3 succeeded in providing elected services during a single appointment 80 to 90 percent of the time. Self-monitoring data also indicated that the awardee engaged C3 patients in evidence-based decision making and adoption of the most effective contraceptive methods, with 50 to 70 percent of all C3 patients selecting intrauterine devices or implants in any given month. Based on Title X data for Missouri in 2017, LARC uptake among Title X patients at C3 was 32.6 percent (more than double the statewide average of 14.6 per cent for all Title X patients) and almost twice the next-highest reported rate for Title X clinics in Missouri (18.0 per cent).

Characteristics of program participants

Washington University's self-reported enrollment statistics show that C3 succeeded in reaching an ethnically diverse patient population that generally reflected the population of the St. Louis metropolitan area: 47 percent of enrollees identified as Black or African American, 35 percent as White, and 11 percent as Hispanic or Latino. Enrolled patients also spanned the reproductive-age spectrum, including adolescence (8 percent were 18 or younger), young adults (39 percent were 19 to 25), and older adults (53 percent were 26 or older). Although nearly half (46 percent) of C3 patients had private insurance, enrollment statistics also indicated that C3 reached lower-income

women. About 20 percent of C3 participants were covered by Medicaid and 31 percent were uninsured at their first clinic visit.

The awardee's self-monitoring data provided additional insight into its effectiveness in enrolling women at risk for unintended pregnancy. As of August 2017, self-reported data at intake from more than 2,300 C3 direct participants indicated that about 42 percent had previously experienced unintended pregnancies. Although about 28 percent of new patients already used LARC methods, about 35 percent of new patients reported using either less-effective contraceptive methods or none at all during their most recent sexual intercourse. About 85 percent of direct participants reported wanting to postpone pregnancy for two years or more, or to avoid pregnancy altogether.

Challenges of estimating program impacts

A rigorous evaluation of C3's impacts on unintended pregnancy rates and associated costs was not possible for three reasons. First, prior experience with the CHOICE program, although critical to the effectiveness of the awardee's implementation of C3, precluded identifying a baseline population lacking exposure to the intervention that could serve as a comparison group for a pre-post evaluation, because women in St. Louis had already been exposed to a similar intervention. Second, although the awardee gathered data on pregnancy intention among C3 patients, no comparable data were available for a comparison group of Medicaid patients (because the claims data do not capture pregnancy intention), making it difficult to determine whether pregnancy outcomes in a comparison group were intended or unintended. Finally, although reduced pregnancy-related costs can be discernible over the shorter term, the larger potential impact of the C3 intervention relates to averted costs associated with childbirth, which would not be observed until one year or more after the delivery of service. The results presented here therefore cannot be interpreted as an evaluation of program impacts.

CONCLUSION

Washington University succeeded in implementing its C3 program to provide comprehensive contraceptive services to women at risk for unintended pregnancy in St. Louis. By the end of the three-year cooperative agreement, the awardee had enrolled 75 percent of its revised enrollment projection, established the C3 clinic, delivered services in a manner consistent with the design proposed in its HCIA R2 application, and successfully adapted the model to address financial barriers to access and potential delays in service delivery related to insurance coverage. The awardee's self-monitoring statistics indicate high levels of patients' uptake of the most effective evidence-based contraceptive methods and high rates of patients' satisfaction with C3's services. Participating clinicians and other staff reported that the program had a positive effect on the delivery of care, providing needed services to women with limited means and limited access to primary or preventive care for whom avoiding unplanned pregnancy was a high priority. Title X data from 2017 further indicated that C3 patients were much more likely to adopt highly effective contraceptive methods than were patients in other Title X-funded family planning

clinics in Missouri. Washington University's continued ability to sustain C3 program services depended on funding and reimbursement. The awardee developed and proposed a payment model that would cover C3 services via a bundled payment for a 90-day episode of contraceptive care, but had not negotiated an agreement with payers by the end of the award period.

PROGRAM SUSTAINABILITY

After its award ended in November 2017, Washington University continued to operate the C3 clinic and sustained two of the program's three components. Although the C3 social worker left, the clinic had access to insurance assistance and social work services in the department's resident-staffed clinic, after moving to a new building. This enabled the awardee to maintain its insurance counseling and patient navigation services.

Washington University also continued providing contraceptive services after the award ended. Despite reducing clinic staff and operating hours after the award ended, the awardee continued to provide timely contraceptive services because the volume of patients referred dropped by about 50 percent. The awardee attributed this drop to it stopping its marketing and outreach efforts, which had encouraged providers and community organizations to refer patients to the program.

Washington University's proposed payment model

Washington University proposed a payment model that would provide a bundled payment for a 90-day episode of contraceptive care. The model used two reimbursement rates, one for LARC provision ($447.46) and one for shorter-acting contraceptive methods ($150.77).

The payment would not have covered the cost of the contraceptive device, but would have covered all other associated costs, including (1) initiation of the contraceptive method, including nonclinician contraceptive counseling, medical intake, clinician services, and pregnancy testing; (2) short-term follow-up and support; (3) facility and administrative charges, including insurance navigation and assistance; and (4) dispensing and insertion fee modifier for LARCs, if applicable. By the end of the award, the awardee had been unsuccessful in negotiating agreements with payers.

Although Washington University continued, temporarily, to provide contraceptive counseling by cross-training research assistants in the department, the awardee reported that it was least likely to be able to sustain this component of C3, which was not reimbursed by any payer. Although the awardee's payment model would have covered the cost of this component, the awardee was unable to reach agreements with the two payers that had shown interest initially, due to both payers changing leadership. The awardee also reported that securing other sources of funding that might have covered nonclinical contraceptive counseling, specifically Title X funding and private donations, was unlikely: Title X, because new federal regulations in summer 2018 threatened the awardee's continuing eligibility for these funds, and private donations because the program no longer had staff with expertise or time to dedicate to fund raising.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: The Wisconsin Department of Health Services

September 2020

John Schurrer, Kristin Andrews Lemos, Linda Barterian, Luke Horner, Bob Schmitz, and Stacy Dale

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

THE WISCONSIN DEPARTMENT OF HEALTH SERVICES

The Wisconsin Department of Health Services (WI DHS) and its partners, the Children's Hospital of Wisconsin (CHW) and the University of Wisconsin Health–American Family Children's Hospital (AFCH), received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the Special Needs Program (SNP). The goal of the program was to address the needs of children with medical complexity (CMC) and their families by providing integrated health care services, including direct and consultative patient care, care management, and care coordination. The target population consisted of CMC with high health care use in the year before enrollment. Specifically, beneficiaries were eligible if they had chronic conditions that affected three or more organ systems requiring ongoing care from three or more specialists and had at least one hospitalization lasting at least five days or 10 or more clinic visits within a 12-month period. The program expanded on the services provided by an existing SNP at CHW. The enhanced SNP began in September 2014. The intervention period funded by HCIA R2 ended in August 2018, after a one-year extension of the award. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- By providing integrated health care, including direct and consultative patient care, care management, and care coordination, the SNP sought to reduce rates of preventable hospitalizations and ED visits, reduce the length of hospital stays, reduce spending associated with ED and hospital use, improve receipt of necessary outpatient services, and improve satisfaction among families and primary care providers.
- Specialists, primary care providers, community programs, and children's caregivers referred children living in the program's catchment areas to the SNP. The SNP teams assessed eligibility of referred children and met with families in person before extending an offer to enroll in the program. About 12 percent of all eligible children (not all of whom were necessarily referred for the program) participated in the SNP.
- Due to selection bias concerns and the inability to identify a comparison group that matched well on key characteristics of participants, it was not possible to conduct a rigorous impact evaluation of this program.

The awardee hypothesized that providing enhanced care management and coordination for CMC would reduce preventable emergency department (ED) and hospital use and spending, improve receipt of necessary outpatient services, and increase satisfaction among families and primary care providers. Physicians and nurse practitioners (NPs) provided direct care to patients during scheduled clinic visits and inpatient stays, oversaw patients' care plans, and collaborated with primary care and specialty providers. Registered nurse and care coordination assistant dyads ensured timely SNP and specialty follow-up appointments. Social workers addressed psychosocial, emotional, and socioeconomic issues that affected care access and adherence.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	The WI DHS implemented the SNP to improve patients' care, reduce preventable ED and hospital use and spending, improve receipt of necessary outpatient services, and improve quality of life for CMC and their families.
Major innovation	Building on an existing program at CHW, the enhanced SNP (1) provided direct and consultative patient care, care management, and care coordination to an expanded group of CMC; (2) implemented the enhanced SNP at a second institution (AFCH); and (3) developed a payment model to sustain the SNP.
Program components	• Direct and consultative patient care • Care management • Outpatient and transitional care coordination
Target population	The awardee sought to engage CMC, regardless of payer, with high health care service use during the year before enrollment. Beneficiaries were eligible for the SNP if they had chronic conditions that affected three or more organ systems requiring ongoing care from three or more specialists and met one of two criteria during the 12 months before enrollment: (1) at least one hospitalization of five or more hospital days or (2) 10 or more outpatient clinic visits. The program was open to Medicaid beneficiaries enrolled in either FFS or managed care.
Total enrollment	The awardee enrolled 685 beneficiaries from September 2014 through August 2017, representing 34 percent of its original three-year goal and an estimated 12 percent of all children potentially eligible for the intervention.[a]
Level of engagement	Nearly all staff survey and clinician survey respondents (96 and 95 percent, respectively) agreed that the enhanced SNP successfully engaged participants. The awardee's self-monitoring data showed an attrition rate of less than 1 percent at both hospitals.
Theory of change or theory of action	The awardee hypothesized that enhanced care management and coordination for CMC would shift the burden of care coordination to the SNP, identify unmet needs, connect families to medical and social resources, and better coordinate care among primary, specialty, and inpatient care. This, in turn, would lead to reduced ED and hospital use and spending and improved satisfaction among families and primary care providers. The SNP providers expected that it would take at least 18 months to see program effects on service use and expenditures.
Award amount	$9,444,864
Effective launch date	September 1, 2014
Program settings	The program operated in two tertiary, acute care children's hospitals (CHW and AFCH) and their associated outpatient clinics.
Market area	The program's catchment area included Milwaukee and Madison, Wisconsin (where CHW and AFCH, respectively, are located), and surrounding regions.
Target outcomes	• Reduce ED visits, hospitalizations, and total hospital days • Decrease total cost of care (9 percent reduction in per participant per month spending) • Improve receipt of necessary outpatient services • Improve participants' quality of life • Improve family and primary care provider satisfaction
Payment model	The awardee took advantage of two new procedural codes that covered the SNP's services not reimbursed under traditional Medicaid FFS.
Sustainability plans and payment model	WI DHS sustained the SNP with a new FFS reimbursement code for targeted case management implemented through a state plan amendment that went into effect in September 2017. WI DHS planned to align the SNP with other care coordination programs and support implementation of the program through a shared savings model in the future.

[a] The analysis estimated the number of eligible children as those with a claim at AFCH or CHW during the intervention period who had three or more CCCs according to Medicaid claims data.

AFCH = University of Wisconsin Health–American Family Children's Hospital; CCC = Chronic Condition Category; CHW = Children's Hospital of Wisconsin; CMC = children with medical complexity; ED = emergency department; FFS = fee-for-service; SNP = Special Needs Program; WI DHS = Wisconsin Department of Health Services.

It was not possible to conduct a rigorous impact evaluation of the SNP because of how the awardee identified and recruited participants into the program. As a result, this report describes only the demographic and health characteristics of Medicaid participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis. Appendix A, Table A.1 describes the identification of the sample used for the descriptive analysis.

Table 2. Key features of the descriptive analysis

Features	Description
Descriptive analysis	Because a credible comparison group could not be identified for the SNP using Medicaid claims data, a low participation rate, and lack of baseline data for children less than 1 year, a rigorous impact evaluation of this program was not possible.[a] As a result, this report describes only the baseline demographic and health characteristics of participants enrolled in Medicaid.
Intervention group for descriptive analysis	The intervention group used in the descriptive analysis included 427 Medicaid beneficiaries who were recruited, screened, and enrolled in the program from September 2014 through August 2017, representing 62 percent of total enrollment during that period. The intervention group for the descriptive analysis did not include 237 Medicaid beneficiaries who had concurrent third-party insurance in the year before enrollment. The analysis excluded another 21 participants because of death within 30 days of enrollment or lack of sufficient enrollment in Medicaid in the year before enrollment.
Limitations	Due to the problems noted above, this report cannot make inferences about the impact of this program on Medicaid costs or other program outcomes.

[a] The analysis estimated the number of eligible children as those with a claim at AFCH or CHW during the intervention period who had three or more CCCs according to Medicaid claims data.

AFCH = University of Wisconsin Health–American Family Children's Hospital; CCC = Chronic Condition Category.

PROGRAM DESIGN AND ADAPTATION

The expanded and enhanced SNP had three key components: (1) direct and consultative patient care, (2) care management, and (3) outpatient and transitional care coordination. The care teams included physicians, nurse practitioners (NPs), registered nurse care coordinators, care coordinator assistants, and social workers.[1]

Direct and consultative patient care

Physicians and NPs provided direct care to patients during scheduled clinic visits with participants, consulted in the inpatient setting, and helped develop and implement patients' care plans. After enrollment, participants typically had office visits scheduled at two months, four months, and every six months thereafter. The care team also provided care as needed when participants experienced medical events that staff at an urgent clinic could manage. In addition, the care team conducted daily rounds if the child was in the hospital.

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of the SNP. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

Care management

Physicians and NPs worked with primary care providers and specialty providers to comanage participants. They collaborated on medical decision making with other physicians and participants' families and provided around-the-clock accessibility by phone to participants' families and other physicians, especially during acute changes in medical conditions and transitions of care between hospital units or from hospital to home.

Outpatient and transitional care coordination

Care coordination teams consisting of a registered nurse and care coordination assistant dyad ensured that participants' SNPs and specialty follow-up appointments occurred in a timely manner; the dyads also served as the primary point of contact for patients' families. Most participants' care plans included a monthly care coordination phone call. In addition to the monthly calls, for hospitalized children, the care coordination team conducted post-discharge follow-up phone calls. Social workers also addressed psychosocial, emotional, and socioeconomic issues that affected access and adherence to care. For example, they helped resolve insurance issues, educated families about government programs, and addressed transportation barriers to care.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

The SNP achieved a high level of engagement with program participants. Nearly all staff survey and clinician survey respondents (96 and 95 percent, respectively) agreed that the enhanced SNP successfully engaged participants. During interviews, care teams noted that monthly follow-up telephone calls helped maintain engagement with patients and families between clinic visits. The awardee's self-monitoring data showed rates of participant withdrawal and loss to follow-up of 1 percent or less at both hospitals throughout the cooperative agreement. More than three-quarters (78 percent) of respondents of the clinician survey indicated that participants' engagement was the most helpful factor in the SNP achieving its goals.

Despite these achievements, the awardee faced several challenges implementing the program. Care teams had expected that after a child enrolled in SNP, the intensity of services required to meet the child's care needs would decrease over time. The teams also planned to classify participants into three tiers of care management and coordination based on need at enrollment. However, the awardee found that participants' health care needs fluctuated over time rather than taper off in a predictable way. Therefore, the awardee dropped the use of tiers because they did not prove useful in predicting changes in participants' needs for SNP services and staffing and caseload demand over time. This made it more challenging than expected for the program to reduce unnecessary service use and spending during the participant's period of enrollment.

In addition, program leaders found that supporting enrolled children with more intense needs than expected stretched program staff capacity. SNP leadership struggled to find qualified candidates for all positions on the care teams. Program leaders and staff attributed recruitment challenges to the unique skill set required for SNP positions, which required strong communication skills, familiarity with the hospital system, a commitment to teamwork, and a passion for the CMC population. Additional challenges in recruiting physicians and NPs included a requirement to work one weekend each month and less competitive salaries compared to other specialties. SNP teams faced challenges training new staff while addressing the needs of an influx of new patients who required more time and attention than existing patients. The care team staff reported that before reaching full staffing in the third year of the program, they did not consistently complete follow-up calls to SNP participants within 72 hours of their discharge from the hospital. Because patients with higher-than-anticipated needs required a higher ratio of staff-to-participants than originally expected, the awardee reduced its enrollment TARGET BY ALMOST HALF, FROM 2,040 TO 1,121.

Implications of program implementation for achieving program goals

- Intensity of service needs did not diminish over time, making it more challenging than expected to reduce unnecessary service use and spending during the participant's period of enrollment
- Supporting enrolled children with more intense needs than expected also stretched program staff capacity and made it difficult to complete intervention protocol.
- Estimating impacts using the participants only would be highly likely to produce biased results, given that parental engagement and interest as well as clinical judgment were key enrollment criteria.
- The low participation rate made it impossible to obtain reliable, unbiased estimates of program impacts by comparing outcomes for all eligibles to outcomes for other children who met the same claims-based eligibility criteria but did not have a claim from a participating hospital.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

Recruiting and enrolling participants

The SNP teams identified eligible children primarily through referrals from specialists, primary care providers, community programs, and children's caregivers. Beginning in the second year of the program, program staff at CHW began to identify and recruit eligible infants receiving care in hospital neonatal intensive care units (NICUs) and pediatric intensive care units (PICUs). AFCH staff also began to identify eligible beneficiaries in inpatient units, but did not recruit patients from the NICU and PICU because their hospital had a separate program serving these populations. At both sites, program teams evaluated each referral to verify alignment with SNP eligibility criteria, and then a care coordination assistant called the family to schedule an assessment office visit before enrolling the child in the program. During these visits, the program

team assessed the patient's needs, parents' engagement and interest in the program, and the patient's potential eligibility for other WI DHS programs. After the assessment visit, program teams invited eligible children to enroll in the SNP, and parents signed a consent form and agreed to participate in the program. Program teams excluded a small number of children whose needs they believed could be better met by a different program—particularly those with co-occurring mental health issues—as well as children of families who did not want to participate in the program. Participants remained enrolled as long as program staff and families agreed that SNP participation continued to offer benefits.

Characteristics of Medicaid participants in the SNP

Because a full year of baseline service use and expenditure data were not available for beneficiaries younger than 1, the analysis presents characteristics separately for those younger and older than 1 (Table 3). Among the 427 Medicaid participants in the study sample, 43 percent were younger than 1 at the time of enrollment. A majority of the beneficiaries younger than 1 were White (56 percent) compared to 37 percent in the group older than 1. A large percentage of participants in both groups resided in metropolitan areas, which reflects the fact that the program operated at hospital sites located in the state's two largest metropolitan areas. On average, both groups had about five medically affected organ systems at the time of enrollment, as measured by the Chronic Condition Category algorithm developed by Feudtner et al. (2014). A substantial proportion of beneficiaries also depended on medical technology, such as a prosthetic heart valve or dialysis. Gastrointestinal disorders were the most common conditions in both groups.

As expected, service use and expenditures during the year before enrollment were substantially larger for SNP participants compared to Medicaid-enrolled children in Wisconsin generally. Total expenditures for SNP participants were $9,617 per beneficiary per *month* compared to a median of $1,762 per beneficiary per *year* in 2014 for all Medicaid-enrolled children in Wisconsin.[2]

Table 3. Baseline characteristics of Medicaid participants

Characteristics	Participants younger than 1 (N = 182)	Participants older than 1 (N = 245)
Demographics		
Age at enrollment, continuous	6.3 months	5.8 years
Male, %	55	57
White, %	56	37
Residence, %		
Metropolitan	91	86
Nonmetropolitan urban	9.2	13
Rural or urban, fewer than 2,500 residents	< 1	1

[2] Data on Medicaid spending in fiscal year 2014 are available from the Kaiser Family Foundation's State Health Facts website. Accessed on February 26, 2020.

Table 3 (*continued*)

Characteristics	Participants younger than 1 (N = 182)	Participants older than 1 (N = 245)
Chronic conditions[a]		
Total number of CCCs present, (continuous)	5.0	4.8
Technology dependent, %	82	73
Cardiovascular condition, %	61	47
Respiratory condition, %	46	28
Gastrointestinal condition, %	79	74
Metabolic condition, %	22	25
Genetic congenital condition, %	42	62
Service use and expenditures during the year before enrollment		
Number of inpatient stays (per 1,000 beneficiaries)	n.a.	1,582
Number of outpatient ED visits (per 1,000 beneficiaries)	n.a.	2,615
Number of clinic visits (per 1,000 beneficiaries)	n.a.	32,063
Total expenditures ($ PBPM)	n.a.	9,617

Sources: Mathematica's analysis of the awardee's finder file and Medicaid claims and encounter data from September 1, 2013 through August 31, 2017.

Note: The baseline period is 365 days before each beneficiary's enrollment date for beneficiaries older than 1. Baseline characteristics for beneficiaries younger than one span the period from birth through program enrollment. The analysis defined the enrollment date as the date of the beneficiary's enrollment in the SNP. It measured all beneficiaries' characteristics during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicaid.

[a] Conditions were grouped using the algorithm developed by Feudtner et al. (2014).

CCC = Chronic Condition Category; ED = emergency department; n.a. = not applicable; PBPM = per beneficiary per month; SNP = Special Needs Program.

Challenges of estimating program impacts

A rigorous impact evaluation was not possible for SNP because of serious concerns about three sources of selection bias. First, SNP staff met with all eligible family members to assess patients' needs and parents' engagement and interest. Some families felt overwhelmed by the severity (and sometimes newness) of the children's condition and chose not to enroll at that time. Thus, program enrollment was greater among children with motivated and engaged parents—an important factor that could not be observed for potential comparison beneficiaries. Second, because the two participating hospitals offered the most advanced care in the state, it was difficult to identify beneficiaries with similar care needs and who did not already receive services from one of the participating hospitals. Third, most participants lived in the Madison and Milwaukee metropolitan areas. Drawing comparison beneficiaries from the nonmetropolitan areas of the state might introduce bias in program impact estimates due to unobservable differences in access to care and other social determinants of health that are associated with the urbanicity of one's residence and can affect outcomes.

About 12 percent of eligible children (as determined by claims data) enrolled in the study. Eligible children included all who had a claim from a participating hospital with a diagnosis

related to three or more body systems using a Chronic Condition Category algorithm developed by Feudtner et al. (2014). This low participation rate made it impossible to obtain reliable, unbiased estimates of program impacts by comparing outcomes for all eligibles to outcomes for other children who met the same claims-based eligibility criteria but did not have a claim from a participating hospital. But estimating impacts using the participants only would be highly likely to produce biased impact estimates, given that parental engagement and interest as well as clinical judgment were key enrollment criteria.

CONCLUSION

Despite challenges reaching its enrollment target, WI DHS successfully engaged program participants and largely delivered services as intended. However, a major medical event triggered enrollment for many participants. As a result, many participants had unexpectedly high health care needs when they entered the program, sometimes requiring many months in the program before their conditions stabilized. Supporting children with more intense needs than expected at enrollment stretched program staff capacity and made it difficult to reduce service use and costs during the program period. Multiple sources of potential selection bias and the associated challenges of finding a comparison group with similar motivational and health characteristics as participants, coupled with the low participation rate, precluded a rigorous evaluation of the program impact on beneficiaries' outcomes.

PROGRAM SUSTAINABILITY

After its award ended in August 2018, WI DHS reported that its partnering hospitals continued the SNP through a state plan amendment that funds SNP activities under a new targeted case management benefit. The state plan amendment became effective in April 2018 and was retroactive to September 1, 2017. The awardee reported CHW and AFCH were the only two hospitals in the state eligible for the benefit at the time. The awardee has since published guidance about program requirements and the targeted case management benefit in an online Medicaid handbook in an effort to help other Wisconsin hospitals become eligible for it. The awardee anticipated that the targeted case management benefit would be

WI DHS proposed payment model

WI DHS was able to continue funding the SNP through a Medicaid state plan amendment. The amendment created two procedural codes that covered the SNP's services not reimbursed under traditional Medicaid fee-for-service. The two procedural codes include:

- Enrollment: a one-time payment of $1,100 for a comprehensive assessment and completion of a patient care plan
- Ongoing care coordination: capitated payments of $450 for each month when the hospital reports at least one care coordination activity for the child

Longer term, WI DHS anticipates aligning the SNP with other care coordination programs and supporting implementation of a shared savings model.

sufficient to sustain the program because it covers the cost of the program for Medicaid beneficiaries, who accounted for about 85 percent of total participants in the SNP.

Longer term, WI DHS anticipates aligning the SNP with other care coordination programs and supporting implementation of a shared savings model. The awardee did not finish developing its alternative payment model with shared savings before the award ended, but intended to continue working on the model with plans to engage commercial payers in the future.

REFERENCE

Feudtner, C., J.A. Feinstein, W. Zhong, M. Hall, and D. Dai. "Pediatric Complex Chronic Conditions Classification System Version 2: Updated for ICD-10 and Complex Medical Technology Dependence and Transplantation." *BMC Pediatrics,* vol. 14, no. 199, 2014. Available at https://link.springer.com/article/10.1186/1471-2431-14-199.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants removed from analytic sample	Number of participants remaining in analytic sample
Total program participants through August 31, 2017		**685**
Not eligible for Medicaid during month of enrollment	12	673
Dually eligible for Medicare and Medicaid in baseline year	1	673
Died within 30 days of enrollment	2	670
Fewer than 30 days Medicaid eligibility in baseline year (if younger than 1 year at enrollment)	1	669
Fewer than 90 days Medicaid eligibility in baseline year (if at least 1 year at enrollment)	5	664
Had third-party insurance coverage during baseline year	237	
Final analytic sample		**427**

Sources: Mathematica's analysis of information from the awardee's finder file and Medicaid claims and enrollment data from September 1, 2014 through August 31, 2017.

SNP = Special Needs Program.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org

Mathematica
Progress Together

Final Report

HCIA Round 2 Evaluation: Yale University

September 2020

Eva Chang, Sonya Streeter, Rumin Sarwar, Kevin Smith, and Nancy McCall

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number/Task Order Number: HHSM-500-2014-00034I/HHSM-500-T0001

Submitted by:

Mathematica
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044

Project Director: Boyd Gilman
Reference Number: 50082

This page has been left blank for double-sided copying.

YALE UNIVERSITY

Yale University received a cooperative agreement under Round 2 of the Health Care Innovation Awards (HCIA R2) to implement the Paramedic Referrals for Increased Independence and Decreased Disability in the Elderly (PRIDE) program. The goal of the program was to use a community-based, short-term care management approach to improve the health of elders and others with impaired mobility in an effort to prevent injuries and associated emergency department (ED) visits and hospitalizations from a fall. The target population consisted of individuals who had fallen or feared falling, and who were residents of the greater New Haven area in Connecticut. The PRIDE program launched in March 25, 2015, and the intervention period funded by HCIA R2 ended on February 28, 2018. Table 1 summarizes the program's key characteristics.

Important issues for understanding the evaluation

- Yale University used a community-based, care management approach to improve the health of elders and others with impaired mobility in an effort to prevent injuries from a fall and associated hospitalizations.
- Yale University's strategy to identify and recruit participants initially focused on patients who had fallen and called 911 for a "lift assist." In response to low enrollment, by the second program year the awardee began recruiting patients in the ED and through self-referrals, and expanded the target population to include individuals who feared falling.
- Due to the use of clinical judgment in recruiting patients in the ED and the inability to replicate the eligibility criteria in claims data, it was not possible to identify a comparison group that was similar to the intervention group. The low participation rate among those eligible according to claims data made it infeasible to obtain reliable unbiased estimates of program impacts using all eligible cases. Thus, no impact estimates are presented in this study.

The awardee hypothesized that providing PRIDE services to individuals who had fallen or feared falling in their homes would reduce falls. Reducing falls would contribute to the goals of the program: reductions in preventable ED visits, hospitalizations, 911 lift-assist calls, and, ultimately, successful reductions in mortality and morbidity in the target population. PRIDE services included in-home assessments related to fall risk, in-home preventive care (such as conducting medication reconciliation), and linkage to primary care (such as scheduling or arranging transportation services to a primary care provider [PCP] visit). PRIDE paramedics and nurses from partnering visiting nurse agencies (VNAs) provided these services. The PRIDE program proposed a payment model under which the regional emergency medical services (EMS) medical director would have received prospective payments for a geographical region to reimburse paramedics, VNA nurses, and transportation providers for providing specific services.

Table 1. Program characteristics at a glance

Program characteristics	Description
Purpose	In the PRIDE program, Yale University served participants who had fallen or were at risk of falling and provided them with in-home interventions and increased linkages to PCPs. Yale aimed to reduce falls and other medical emergencies that contribute to preventable ED visits, hospitalizations, and 911 calls.
Major innovation	The major innovation of the PRIDE program was employing paramedics and VNA nurses to visit participants in their homes to identify their medical and social needs and connect them to appropriate resources.
Program components	Short-term care management
Target population	The awardee sought to engage individuals in the greater New Haven area of Connecticut who had fallen or who were at risk of falling.
Participating providers	Paramedics and VNA nurses employed by the PRIDE program.
Total enrollment	The awardee enrolled 5,222 unique participants (109 percent of its original enrollment target) across 3 recruitment arms: after a 911 call, during an ED visit, and from self-referral. Of those enrolled, 2 percent were enrolled after a 911 call, 50 percent during an ED visit, and 48 percent from self-referrals.
Level of engagement	Of the 5,222 unique participants, 66 percent received a visit from both a paramedic and a VNA nurse, 7 percent received either a paramedic or VNA visit, and 27 percent did not receive either type of visit. The proportion of participants who received paramedic and VNA home visits varied across recruitment arms.
Theory of change or theory of action	If a paramedic and VNA nurse visited participants in their homes to perform fall risk assessments, deliver preventive care, and link participants to primary care, the number of falls and other medical emergencies will be reduced. This would, in turn, lead to reductions in preventable ED visits, hospitalizations, 911 lift-assist calls, and, total health care spending.
Award amount	$7,159,976
Effective launch date	The program began operating on March 25, 2015.
Program settings	Participants' homes
Market area	Urban and suburban areas
Market location	New Haven County, Connecticut
Target outcomes	• Reduction in lift-assist calls • Preventable ED visits • Hospital admissions • Total expenditures • Mortality
Payment model	Prospective, population-based payment for care management and coordination services
Sustainability plans	Yale University continued a modified version of the PRIDE Program funded by the Yale–New Haven Hospital System. The modified program focuses on recruiting potential participants who are admitted to the ED or inpatient care.

ED = emergency department; PCP = primary care provider; PRIDE = Paramedic Referrals for Increased Independence and Decreased Disability in the Elderly (PRIDE) program; VNA = visiting nurse agency.

It was not possible to conduct a rigorous impact evaluation of the PRIDE program because of the way in which the awardee identified and recruited participants. As a result, this report describes only the demographic and health characteristics of Medicare FFS participants, and does not present estimates of program impacts. Table 2 summarizes the key features of the descriptive analysis.

Table 2. Key features of the evaluation

Features	Description
Descriptive analysis	Due to how the program identified and recruited participants, it was not possible to conduct a rigorous impact evaluation of this program. Identifying comparison beneficiaries similar to the intervention participants recruited from the ED could not be replicated in health care claims because hospital staff used clinical judgment to identify and recruit participants. Using a broader treatment group sample of individuals meeting the eligibility criteria according to claims to produce unbiased impact estimates would have had very low power to detect even large impacts, due to the low participation rate.
Intervention group for descriptive analysis	The intervention group for the descriptive baseline analysis was based on 15 percent of total unique study participants. Study participants were defined by having signed a consent form in the ED from March 2015 through February 2018. Among the 2,670 unique beneficiaries the awardee identified as an ED participant, 326 participants either were not enrolled in Medicare or could not be identified in the Medicare enrollment database, and 915 ED participants did not meet the claims-based inclusion criteria of the study, such as being enrolled in Medicare Parts A and B and having Medicare FFS as the primary payer. The study sample was further refined to match the initial program enrollment criteria for ED patients to focus on patients who were discharged to home or home care. An additional 646 participants were excluded from the study sample if the participant did not have an outpatient ED claim starting on or within 3 days before the enrollment date or died within 30 days of the enrollment date. The final size of the group for the descriptive analysis was 782 PRIDE participants.
Limitations	Due to the problems noted above, no inferences can be made about the impact of this program on Medicare costs or other program outcomes.

FFS = fee-for-service.

PROGRAM DESIGN AND ADAPTATION

Yale University's service delivery model focused on short-term care management in which participants received a home visit by a PRIDE paramedic and a home visit by a PRIDE VNA nurse.[1] After a participant enrolled in the program, a PRIDE paramedic completed an initial in-home assessment, which considered the participant's health status and residential safety. At the end of the assessment, and if the participant agreed, the PRIDE paramedic scheduled an appointment for the participant with his or her PCP. If the participant desired transportation assistance, the paramedic arranged for transportation to and from one PCP visit. A PRIDE VNA nurse then completed a second in-home visit, which involved a participant health assessment and preventive care measures (for example, appropriate configuration of a walker). While in the home, the VNA nurse assessed the individual's overall health, focusing on the need for ongoing

[1] The Third Annual Evaluation Report provides additional details on the design and implementation of Yale University's PRIDE program. It is available at https://downloads.cms.gov/files/cmmi/hcia2-yr3evalrpt.pdf.

nursing care, physical and occupational therapy, or durable medical equipment; assessing for suspected neurological conditions; assessing mobility; and performing a medication review and reconciliation. The VNA nurse provided the participant with resources and services available within the community and how to access them, and educated the participant and household members on strategies to prevent falls. The VNA nurse communicated the assessment findings to the participant's PCP, depending on the participant's assessment results.

Uptake of the paramedic and VNA home visits was low in the first two program years, so the awardee introduced changes to better engage patients. In the third program year, Yale set goals for completing the paramedic (initial) home visit within 14 days of enrollment and the VNA (second) home visit within 30 days of enrollment. There was no defined time frame for completing either visit during the first two program years. Also in the third program year, Yale adapted the program to offer a combined paramedic and VNA nurse visit if a participant required translation services or lived in a neighborhood perceived by staff to be unsafe.

ACHIEVEMENTS AND CHALLENGES OF PROGRAM IMPLEMENTATION

Awardee staff reported experiencing challenges in recruiting and enrolling patients, in and engaging patients, and in staff changes early in the program.

Recruiting and enrolling participants

Early in the program, Yale University recognized that the program's initial participant identification and recruitment focus was too narrow and contributed to low enrollment, so the awardee expanded the target population and recruitment strategy. The awardee originally only recruited patients who had fallen and called 911 for a lift assist but were determined to be uninjured and not transported for further medical attention. By the start of the second program year, Yale University expanded the participant criteria to include individuals who had fallen and those who were concerned about a fall, even if they had never experienced one. To recruit patients from the expanded target population, Yale University continued enrolling participants after a 911 lift-assist call, but added two new recruitment mechanisms: patients in the ED and self-referred patients. Patients enrolled through the ED arm included those presenting who had fallen or who feared falling and were treated in the ED and discharged home or admitted to the hospital. (Participants were considered enrolled upon signing a program consent form.) Based on the success of the two new recruitment mechanisms, Yale University increased its efforts using these recruitment strategies in the second and third program years by adding a second ED and by expanding its catchment area for community recruiting events. By the end of the cooperative agreement, Yale University surpassed its original enrollment target. ED visits and self-referrals made up 50 and 48 percent of total enrollment, respectively, and the original 911 lift-assist call made up 2 percent of total enrollment.

Engaging participants

Over the three-year cooperative agreement, 66 percent of participants received both a paramedic and a VNA home visit, 7 percent received either a paramedic or VNA visit, and 27 percent did not receive any home visit. However, the proportion of participants who did not receive paramedic and VNA home visits varied across recruitment arms. ED participants were less likely to have received both paramedic and VNA visits (51 percent of ED participants received both visits) compared to self-referral participants (81 percent) and 911 lift assist participants (69 percent).

The awardee struggled to establish connections between participants and PCPs. PRIDE paramedics found that participants frequently declined these services, saying that they already had appointments with their PCPs or that a family member would provide transportation. Self-monitoring data provided by the awardee indicated that one in five participants accepted PRIDE paramedics' offers to schedule PCP appointments; only 3 percent of those who accepted this service requested transportation to a PCP appointment.

Implications of program implementation for achieving program goals

- To address low enrollment, the awardee expanded their target population and recruitment strategy early in the program. In addition to enrolling participants after a 911 lift-assist call, Yale also recruited participants while they were in the ED and from self-referral.
- Challenges in recruiting and hiring limited the program's ability to engage participants early in the program. Engagement was particularly low among ED participants.
- During the last two program years, the program hired a variety of staff to boost enrollment and patient engagement rates.

Staffing changes

Yale University overcame several challenges in recruiting, hiring, and retaining appropriate program staff to deliver PRIDE services. Yale University hired a variety of staff to boost enrollment and patient engagement rates. To improve enrollment and engagement, the awardee hired a central office staffer during the second year to manage paramedics' schedules, organize community outreach efforts, and oversee the PRIDE ED staff. In the second program year, Yale University hired new PRIDE ED staff to expand PRIDE recruitment to a second ED (Yale New Haven Hospital's St. Raphael's campus). Unlike previous PRIDE ED staff, the new hires were nursing and public health graduate students who had experience with health care, research studies, and use of Epic (the hospital system's electronic medical record system). Finally, delays in hiring and bringing paramedics on board limited the program's capacity for providing care management services in the first and second program years; the program altered its delivery model by hiring a full-time nurse and full-time paramedic in the third year.

DESCRIPTIVE ANALYSIS OF PARTICIPANT CHARACTERISTICS

The descriptive analysis is based on 782 Medicare fee-for-service (FFS) beneficiaries who had an outpatient ED visit and were discharged home from March 25, 2015, to February 28, 2018, and who agreed to participate in the PRIDE program (representing 15 percent of total enrollment) at 2 EDs in the Yale-New Haven Health System. The beneficiaries in the descriptive analysis did not include all 5,222 participants who enrolled in the Yale PRIDE program. (Appendix A, Table A.1 provides detail on the identification of the analytic sample.) The descriptive analysis focuses on the ED participants because these participants had a qualifying medical event that could be identified in Medicare claims data. Because participants could enroll in the program multiple times through multiple recruitment arms, those who enrolled through the ED and another arm were eligible to be included in the study sample at the time of their qualifying ED visit.

An analysis of Medicare FFS claims from the participating EDs indicated that the ED participants represented only 3 percent of all other beneficiaries presenting to participating EDs with diagnostic information in Medicare claims data indicating or being suggestive of having had a fall. Secondary ICD-9 E-codes (that describe the external cause of injuries) were used to identify beneficiaries that presented to the ED with a diagnosis indicating that the beneficiary had fallen. Suggestive diagnoses included syncope, dizziness, vertigo, and fractures. Participants also appeared sicker than beneficiaries who presented EDs with the same clinical indicators of a fall but were not enrolled in the PRIDE program. These beneficiaries might not have been approached or might have declined to participate when recruited by PRIDE staff in the ED. Thus, the use of clinical judgment in recruiting patients in the ED and patient refusal would lead to biased impact results. The participation rate among claims-based eligible patients was too low for an analysis of all eligibles to detect even large impacts on participants.

Characteristics of Medicare FFS ED participants in the PRIDE Program

Among the 782 ED participants enrolled in Medicare FFS, the average age was 79 years (Table 3). More than 70 percent of the Medicare FFS participants were White and female. Eighty percent of participants were eligible for Medicare because of age and almost 40 percent of participants were dually eligible for Medicare and Medicaid. The average hierarchical condition category (HCC) risk score of participants was 2.1, indicating that their expected Medicare annual spending was at least twice as high as the average for Medicare FFS beneficiaries nationally. During the baseline year, these participants had 615 acute care hospitalizations and 2,333 ED visits per 1,000 Medicare FFS participants per year.[2] Average Medicare spending per beneficiary per month (PBPM) was $2,229 during the baseline year.

[2] The number of ED visits was top-coded at the 98th percentile of the intervention group's distribution of ED visits during the baseline period to reduce the influence of outliers.

Table 3. Baseline characteristics of Medicare FFS participants

Characteristics	ED Participants (N = 782)
Demographics	
Age at enrollment, years	79
Age group, %	
Younger than 65	6.1
65 to 74	25
75 to 84	35
85 and older	34
Female, %	70
White, %	71
Original reason for Medicare eligibility, %	
Age as original reason for Medicare eligibility	80
Medicare-Medicaid dual status, %	
Medicare-Medicaid dual status	39
HCC score[a]	
Mean	2.1
25th percentile	1.1
Median	1.7
75th percentile	2.6
Service use and expenditures during the year before enrollment	
Any hospitalizations, %	34
Hospital stays (per 1,000 beneficiaries)	615
ED visits (per 1,000 beneficiaries)[b]	2,333
Total Medicare expenditures ($ PBPM)	2,229

Source: Mathematica's analysis of information from awardee's finder file and Medicare claims and enrollment data from March 2014 through February 2018.

Note: The baseline year is defined as the 365 days before and including each beneficiary's enrollment date. The enrollment date is defined as the date on which the participant signs a consent form in the ED. All beneficiary characteristics were measured during or as of the end of the baseline year.

The statistics are weighted means, with participant weights proportional to the number of months during the 12-month baseline period that the participant was enrolled in Medicare.

None of the differences between treatment and comparison groups in any of the baseline characteristics differed statistically from zero at the 0.10 level, two-tailed test, except for the ED visits per 1,000 beneficiaries measure ($p < 0.01$). ED visit measures include observation stays.

[a] The HCC score incorporates diagnosis history and demographics to estimate a score representing the expected costs of a Medicare beneficiary in the upcoming year. A score of one represents average expected expenditures. HCC scores were calculated by using the most recently available HCC algorithms.

[b] The number of ED visits was top-coded at the 98th percentile of the intervention group's distribution of ED visits during the baseline period to reduce the influence of outliers.

ED = emergency department; HCC = hierarchical condition category PBPM = per beneficiary per month.

Challenges of estimating program impacts

A rigorous impact evaluation was not possible for this awardee because of how the program identified and recruited participants. The eligibility criteria for participants who enrolled through self-referral or through 911 lift-assist requests could not be replicated in Medicare or Medicaid claims because they did not have a qualifying medical event. For ED participants, the program could not identify comparison beneficiaries similar to the intervention participants in health care claims because PRIDE and hospital staff used clinical judgment to recruit participants. Evidence indicates that participants were sicker than other beneficiaries presenting to participating EDs who appear to be eligible based on Medicare claims and enrollment data. These beneficiaries might not have been approached by PRIDE staff or might have declined to participate when recruited by PRIDE staff in the ED. Because of the lack of a qualifying event for self-referral and 911 lift-assist participants and the use of clinical judgment in recruiting patients in the ED, Medicare claims data could not be used to identify comparison beneficiaries similar to program participants. Furthermore, the participation rate among claims-based eligible patients was too low for an analysis using all eligibles to detect even large impacts on participants. Thus, no impact estimates are presented in this study.

CONCLUSION

Yale University was partially successful in implementing the PRIDE program that used a community-based, short-term care management approach to improve the health of elders and others with impaired mobility among residents of the greater New Haven area in Connecticut. Implementation challenges included early low enrollment numbers that necessitated the expansion of the target population and recruitment strategy, low engagement with PRIDE services (paramedic and VNA home visits and PCP appointment scheduling), and difficulties in recruiting, hiring, and retaining appropriate program staff to deliver PRIDE services. Because of the lack of a qualifying medical event for self-referral and 911 lift-assist participants and the use of clinical judgment in recruiting patients in the ED, Medicare claims data could not be used to identify comparison beneficiaries similar to program participants to conduct a rigorous impact evaluation.

PROGRAM SUSTAINABILITY

After the award ended in August 2018, the Yale-New Haven Hospital System agreed to fund a modified version of the PRIDE program. Changes to the program included restricting program enrollment only to patients treated in the ED or hospital; combining paramedics and VNA nurses into one visit, instead of two separate visits; and requiring that the home visit is completed within 48 hours after the participant is discharged from the hospital or ED.

Yale University proposed a payment model in which the regional emergency medical service medical director would receive a prospective, population-based payment for all beneficiaries in a geographic region to proactively address fall risk for individuals who had already fallen or were

at risk of falling in their homes. By the end of the award, the awardee had planned to, but ultimately did not, calculate payment amounts or secure interest from any payers. Commercial payers the awardee contacted were uninterested, and Medicare reimbursements for program services would require a change in payment policy, which currently pays emergency medical responders only for treatment provided during transport. Given these challenges, Yale University did not continue to develop or implement its payment model.

Yale's proposed payment model

Yale University proposed a payment model in which the regional EMS medical director would receive a prospective, population-based payment for all beneficiaries in a geographic region to proactively address fall risk for individuals who had already fallen or were at risk of falling in their homes. The EMS medical director would have served as an intermediary who reimbursed paramedics, VNA nurses, and transportation providers for providing the program services: an initial in-home assessment by the responding paramedic, an initial home visit by a VNA nurse, and a scheduled round-trip transportation to the participant's PCP.

This page has been left blank for double-sided copying.

Appendix A

Identifying sample for descriptive analysis

This page has been left blank for double-sided copying.

Table A.1. Identification of final sample for descriptive analysis

	Number of participants excluded from analytic sample	Number of participants remaining in analytic sample
Total program participants[a]		**5,222**
Number of participants enrolled through self-referral or 911 lift assist alone	2,552	2,670
Participants not enrolled in Medicare or could not be identified in Medicare enrollment files	326	2,344
Participants who did not meet the standard claims-based inclusion criteria		
Not enrolled in both Medicare Parts A and B	136	2,208
Enrolled in Medicare Advantage	747	1,461
Medicare was not the primary payer	11	1,450
Lacked 90 days of FFS enrollment during baseline period	21	1,429
Participants who did not meet the qualifying ED visit inclusion criteria		
Enrolled after the cutoff	1	1,428
Did not have claims for an outpatient ED visit within 0 to 3 days of the enrollment date and was not discharged to home, self-care, or home care	643	785
Died within 30 days of enrollment	3	782
Final analytic sample		**782**

Source: Mathematica's analysis of information from the awardee's program finder file and Medicare claims and enrollment data from January 1, 2012 through August 31, 2018.

[a] The finder file included 5,665 observations. Observations were de-duplicated within each recruitment arm and the first enrollment within each recruitment was retained for analysis.

ED = emergency department; FFS = fee-for-service.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org