MATHEMATICA Policy Research

REPORT

VOLUME II: Individual Program Summaries

Evaluation of Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report

March 2016

Lorenzo Moreno
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sheila Hoag
Linda Barterian
Laura Blue
Katharine Bradley
Emily Ehrlich
Kristin Geonnotti
Lauren Hula
Keith Kranker
Rumin Sarwar
Rachel Shapiro
KeriAnn Wells
Joseph Zickafoose
Sandi Nelson
Kate Stewart
Frank Yoon

With the following teams: Impact, Implementation, Data Processing, Survey, Statistics, and Editorial and Production Coordination

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850
Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Atlantic General Hospital

March 2016

Linda Barterian
Keith Kranker
Rumin Sarwar
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sandi Nelson
Laura Blue
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850
Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Lorenzo Moreno

Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

ATLANTIC GENERAL HOSPITAL

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by Atlantic General Hospital (AGH) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the AGH program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on participants' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF AGH

AGH received a three-year, $1.1 million dollar HCIA to implement a patient-centered medical home (PCMH) model in all seven of its primary care practices located throughout eastern Maryland and southern Delaware. Table I.1 summarizes key features of the program. The AGH PCMH program included two key components: (1) care coordination for participants diagnosed with chronic conditions and (2) post-hospitalization care transitions support for participants discharged from AGH with any diagnosis. Through a partnership with the Worchester County Health Department (WCHD), the program also provided participants with assistance overcoming social and financial barriers to self-care. In addition, AGH used health information technology (health IT) and conducted community education and outreach to support the PCMH model. As part of this effort, AGH developed a patient portal, designed to enable participants to communicate directly with providers, request appointments and referrals, order prescription refills, and access their medical records and health information. AGH also partnered with 15 faith-based community organizations to disseminate information on PCMH services and provide on-site access to the portal. By the end of the award in June 2015, AGH aimed to reduce hospital admissions and emergency department (ED) visits by 20.0 percent and total cost of care by 15.5 percent.

Table I.1. Summary of AGH PCR program

Program name	Atlantic General Hospital
Award amount	$1,097,512
Implementation date	January 1, 2013
Award end date	June 30, 2015
Program description	• Implement a PCMH model at AGH and its seven primary care practices, including care coordination for patients with chronic conditions and care transitions support for all patients discharged from AGH • Use health IT and conduct community education and outreach to support the PCMH model by improving participants' awareness of support services and increasing access to health information
Innovation components	Care coordination, care management, care transitions, patient-centered care, health IT, risk-stratification
Intervention focus	Practice

Table I.1 *(continued)*

Workforce development	Care coordinators, social worker, registered nurse, PCMH coordinator, health promotion/data specialist
Target population	Frequent users of inpatient services; patients with chronic conditions
Program setting	Provider-based (hospital and primary care practices)
Market area	Regional (eastern Maryland and southern Delaware)
Market location	Rural (Worchester County, a federally designated medically underserved area)
Core outcomes	• 20.0 percent reduction in hospital ED visits • 20.0 percent reduction in hospital admissions • 15.5 percent reduction in total cost of care

Source: Review of AGH program reports, June 2015.

Note: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health IT systems, and other operational activities.

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how it was adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of AGH's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in April 2014 and April 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

The AGH PCMH program included two key components: (1) care coordination for participants diagnosed with chronic conditions and (2) care transitions support for participants discharged from AGH with any diagnosis (Table II.A.1). During the first year of implementation, AGH added a third component, the Keeping in Touch (KIT) program, to meet the needs of participants who did not require the level of intervention services offered through its care coordination and care transitions programs but who could benefit from less intensive follow-up to help manage their conditions.

Table II.A.1. Key details about program design and adaptation

	Program component		
	Care coordination	**Care transitions**	**KIT**
Target population	Medicare beneficiaries with a primary diagnosis of COPD, CHF, or DM; expanded to others expected to benefit, such as those with other chronic conditions (for example, obesity or hypertension), or social needs or mental health issues who required assistance to adhere with medication regiments and care plans, even if non-Medicare or younger than 65.	All patients with an AGH PCP who were discharged from AGH with any diagnosis.	Patients discharged from care transitions or care coordination programs who required less intensive follow-up support to manage their conditions.
Patient identification	Providers identified patients in their panels who met target population criteria. Program staff used hospital discharge data to identify frequent users, defined as more than 2 admissions or ED visits within 6 months, and potential frequent users, defined as patients with 3 or more chronic conditions who do not meet the frequent user definition and notified providers of patients meeting either criteria.	Program staff used hospital discharge data to identify patients. A care coordinator also reviewed discharge summaries daily to identify patients with elevated risk for readmission using the LACE index, which predicts a patient's readmission risk based on length of stay, acute admission through the ED, comorbidities, and ED visits during the past six months (Van Walraven et al. 2010) for prioritized recruitment.	KIT nurses notified providers of discharged patients with issues that might impair effective self-care.
Patient recruitment and enrollment	Providers recruited patients face to face during office visits, explaining benefits and encouraging participation; patients who agreed to participate were referred through EHR to the care coordinator. Care coordinators made a brief 5- to 10-minute introductory call to the participant to confirm participation, describe the program, and schedule the first follow-up call.	A care coordinator visited the patient in the hospital to introduce the program; an informational brochure was mailed to the patient's house before discharge. The care coordinator made a follow-up call within 72 hours of discharge to explain the program, answer questions, and enroll the patient in the program.	Providers discussed discharge plans with participants and made referrals based on an assessment of ongoing support needs.
Service delivery protocol	After enrolling a participant, care coordinators conducted a 30-minute call with the participant, during which they reviewed conditions, assessments, goals, and a care plan. Thereafter, care coordinators reviewed the participant's progress by monitoring lab results, attending the participant's office visits, and through weekly calls with participants, increasing frequency to 2 to 3 times a week for those with unstable conditions. Participants were discharged after meeting care plan goals, typically within 6 to 12 months.	After a patient agreed to enroll, a care coordinator scheduled a call to assess transition needs and schedule participant follow-up appointments with providers; typically, conducted weekly calls during 30 days post-discharge, increasing frequency for participants with unstable conditions. Those identified as needing additional assistance in the home were referred for home visits. All participants were discharged after 30 days; providers of those at high risk for readmission after 30 days were notified by the care coordinator.	KIT nurses made brief weekly calls to participants to identify any emerging concerns and notified care coordinators and providers of any issues with participants' self-care.

Table II.A.1 *(continued)*

	Program component		
	Care coordination	Care transitions	KIT
Adaptations	Yes; AGH planned to include only patients with COPD, CHF, and DM. From program inception, the target population expanded when providers identified and referred patients with other conditions and needs who could benefit from the program.	Yes; during the last year, the care transitions nurse developed a relationship with an area nursing home and began participating in patient rounds to monitor participants who transferred to these facilities during the 30 days following discharge.	No

Sources: Interviews from second site visit, April 2015; document review, June 2015; Van Walraven et al. 2010.

[a] Reportedly, this call often occurred within 24 hours of discharge.

COPD = chronic obstructive pulmonary disease; CHF = congestive heart failure; DM = diabetes mellitus; EHR = electronic health record; PCP = primary care provider.

b. Target populations and patient identification, recruitment, and enrollment

Table II.A.1 provides key details about the target populations and the patient identification, recruitment, and enrollment processes for each component. AGH initially designed the care coordination program to target Medicare beneficiaries with a diagnosis of chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), and diabetes mellitus (DM). From inception, however, AGH expanded the program to include patients with other conditions and needs who referring providers believed could also benefit from participation in the program. The care transitions component included all patients who had an AGH primary care provider and were discharged from AGH with any diagnosis. Providers referred patients discharged from either the care coordination or care transitions program to the KIT program if the provider identified a need for ongoing but less intensive follow-up to help the patient continue to manage his or her conditions effectively.

c. Service delivery protocols

Initially, AGH planned to develop detailed patient protocols for the care coordination and care transition program components. However, care coordinators believed the proposed intervention protocol was inflexible and would limit their ability to customize participants' care plans and follow-up. Therefore, AGH began developing a more flexible approach using disease-specific clinical guidelines and planned to integrate these guidelines into the PCMH. Table II.A.1 provides key details about the service delivery protocols for each component of the program.

d. Intervention staff and workforce development

AGH created several new core clinical staff positions to support implementation of the care coordination and care transitions program components across all seven AGH primary care practices (Table II.A.2). The program's three care coordinators were registered nurses with extensive clinical experience in inpatient and outpatient settings as well as experience using AGH's EHR. The care coordinators divided responsibilities with each other to serve as a primary point of contact for patients participating in the program. In addition, AGH used HCIA funds to support a nurse and social worker from the WCHD who conducted patient needs assessments

and home visits as requested by providers. Two retired nurse volunteers staffed the KIT program. Finally, during the last year of the program, AGH added two administrative positions to support the PCMH: a program manager to supervise day-to-day operations and a data specialist to manage data collection and reporting.

Table II.A.2. Key details about intervention staff

	Staff responsibilities by program component		
	Care coordination	**Care transitions**	**KIT**
Care coordinator (RN) (3 FTEs across programs)	Served as the main point of contact for participants enrolled in the program, provided telephone follow-up, attended participants' office visits, and consulted with providers to help participants manage their conditions (2 FTEs)	Assessed participants care transitions needs, provided telephone follow-up, made participants' post-discharge follow-up appointments (position not supported by the HCIA)[b] (1 FTE)	--
WCHD nurse (RN (1 FTE)	Assisted with participants' needs assessments, conducted in-home visits as requested by providers	Assisted with participants' needs assessments, conducted in-home visits as requested by providers	--
WCHD social worker (RN) 0.5 FTE)	Assisted with participants' needs assessments, conducted in-home visits as requested by providers	Assisted with participants' needs assessments, conducted in-home visits as requested by providers	--
Retired nurse (RN) (2 part-time volunteers)	--	--	Made follow-up calls to participants, alerted care coordinators and providers of any emerging participant concerns
Program manager (RN)[a] (1 FTE)	Supervised day-to-day program operations	Supervised day-to-day program operations	Supervised day-to-day program operations
Data specialist[a] (1 FTE)	Completed mandatory program reporting, conducted high-risk surveillance, monitored outcomes	Completed mandatory program reporting, monitored outcomes	--

Sources: Interviews from second site visit, April 2015, and document review, June 2015.

Note: A solo practice practitioner joined AGH and participated in the PCMH during the final program year.

[a] Originally, the AGH clinical director (a position not supported by the HCIA) managed day-to-day operations and program data management and reporting. In the last program year, AGH used HCIA funding to add the program manager position to take over day-to-day program management and the data specialist position to take over data collection and reporting.

[b] As of June 2015, AGH reported that the organization spent $82,956 for in-kind expenditures for staffing the program.

FTE = full-time equivalent; RN = registered nurse.

At the beginning of its program, AGH conducted training for staff, providers, and partners, including education on National Committee for Quality Assurance (NCQA) PCMH standards, the PCMH philosophy of care, and health literacy. In addition, KIT volunteer nurses and WCHD

program staff received training in use of the AGH EHR. During the second year of implementation, care coordinators completed a course in motivational interviewing focused on participant engagement strategies. AGH did not identify any additional staff education needs and did not conduct any other program training.

2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators and self-reported information included in AGH's quarterly self-monitoring and measurement reports. Table II.A.3 summarizes AGH's self-reported program implementation measure targets and achievements.

a. Program enrollment

AGH exceeded its goal to enroll 20 percent of the 1,314 projected total number of Medicare beneficiaries with a primary diagnosis of COPD, CHF, or DM in the PCMH program, equivalent to a total of 263 participants. As of June 2015, AGH enrolled 1,460 participants in the PCMH program. Although AGH expanded the target population for the care coordination program to include patients with other conditions who providers believed could benefit from care coordination services, most patients referred by providers had at least one of the original targeted conditions of CHF, COPD, or DM. Further, the proportion of patients with a diagnosis of CHF, COPD, or DM who participated in the program increased from about 50 percent initially to 90 percent in the last year of the program. (AGH did not report program enrollment by individual diagnosis, and the total number of patients with a diagnosis of CHF, COPD, or DM seen by AGH providers was not available.)

AGH reported high rates of participation among patients referred to the care coordination program by providers, but did not meet its goal of 1 percent of participants opting out. The percentage of participants opting out of the program after enrollment averaged 9 percent throughout the award period but improved over time and remained below 5 percent during the last six months of the program. This could reflect improved provider effectiveness both in identifying patients who could benefit from program services and were ready to make changes to improve their health, as well as in convincing patients to participate. It also provides some evidence for increased care coordinator effectiveness in engaging participants after receiving training in motivational interviewing.

b. Service measures

AGH met its only service measure goal: care coordinators contacted every participant referred to the PCMH program. Contact included cases in which care coordinators left messages but were unable to speak directly with the participant. After three attempts, participants who did not return messages were discharged from the PCMH. The number of participants discharged after three failed attempts was not tracked.

AGH program administrators and staff reported that they faithfully adhered to the PCMH program model in delivering care coordination and care transitions services, but lack of defined

targets for delivery of services for individual program components limited our ability to assess program implementation effectiveness. Reported participant encounter information included only total volume and estimates of the proportion of follow-up conducted by telephone and in person during participant enrollment. As of June 2015, PCMH care coordinators had a total of 7,422 encounters with program participants. Program administrators reported that about 80 percent of encounters consisted of follow-up telephone calls; the remaining 20 percent consisted of in-person encounters during participants' office visits at provider practices. In addition, the WCHD nurse and social worker together conducted five to seven home visits per week, averaging about 90 minutes per visit.

Table II.A.3. AGH self-reported program implementation measures

Measure	Target	Actual	Met target?	Adaptation?
Program enrollment	263	1,460 PCMH participants, all program components combined	Yes	Yes, expanded target population for care coordination program to include patients with other chronic conditions or social or mental health needs who providers believed would benefit from the program
Opt-out rate	1 percent	9 percent	No	Yes; program staff completed a course in motivational interviewing which provided guidance in how to engage patients
Percentage of PCMH participants contacted by care coordinators	100 percent	100 percent[a]	Yes	--
Participant encounters	Not specified	7,422	--	--
Program staffing	4.5 FTEs (year 1)	6.5 FTEs (year 3)	Yes	Yes; after the first year, AGH hired a program manager and a data specialist to improve the PCMH program
Average care coordinator caseload	50 patients	30 to 42 patients	No	No; but lower caseloads enabled care coordinators to adjust the intensity of support services to accommodate needs of participants with complex conditions
Training	Varied by component	Varied by component	Yes	Yes; program staff completed a course in motivational interviewing that provided guidance in how to engage patients and provide patient-centered care

Sources: Interviews from second site visit, April 2015, and document review, June 2015.

[a] Care coordinators contacted every PCMH patient. Contact included some cases in which care coordinators left messages but were unable to speak directly with the patient. After three attempts, patients who did not return messages were discharged from the PCMH.

AGH did not assess achievement of care plan goals among patients discharged from the program. Instead, it reported the percentage of PCMH participants who either achieved or are actively working toward achieving care plan goals. Because this measure is difficult to interpret, it is not included in our assessment of implementation effectiveness. Notably, AGH administrators recognized the limitation of this measure and, although it remained part of their routine measuring and monitoring report, they did not rely on it as a valid measure of program performance.

AGH informally monitored education and outreach efforts to support the PCMH but did not set targets for these activities. Program administrators reported that a review of computer usage logs showed that the use of computers at faith-based partner organizations by members increased over time, and the proportion of searches conducted on health topics also increased. AGH did not assess usage of the patient portal to support the PCMH, nor was data available for us to assess its use.

c. Staffing measures

AGH exceeded its original program staffing target of 4.5 full-time equivalents (FTEs). After the first year of implementation, program administrators identified a need for additional administrative support to continue to improve their PCMH program. During the last year, they increased program staff to 6.5 FTEs with the addition of two new positions to supervise day-to-day operations and manage data collection and reporting.

Average care coordinator caseloads remained lower than expected throughout the program, ranging from 30 to 42, compared with AGH's target of 50 participants per coordinator. However, staff reported that these lower caseloads enabled care coordinators to adjust the intensity of support services they delivered to meet the needs of high-risk participants with complex conditions, such as increasing the frequency of follow-up calls to participants. Unstable participants with high needs reportedly represented about one-fourth of care coordinators' caseloads.

AGH had very high retention of PCMH program staff. AGH itself maintained all of its program staff, but experienced some turnover in WCHD staff committed to the program. After the first year, WCHD reassigned the full-time nurse and part-time social worker staff supporting the AGH PCMH to other programs at WCHD in response to internal project needs. Then, WCHD replaced the full-time nurse with a part-time nurse, resulting in a reduction in staff supporting the PCMH. The AGH care transitions coordinator took on an increased caseload, and the replacement part-time WCHD social worker increased her level of effort to maintain support services with existing staff positions.

At the beginning of its program, AGH conducted PCMH training for staff, providers, and partners, including education on National Committee for Quality Assurance PCMH standards (four-hour course, 5 trainees); the PCMH philosophy of care (four-hour course, 255 trainees, including the AGH board of directors and providers, hospital staff, and county aging conference attendees); and health literacy (one-hour course, 52 trainees) as planned. During the second year of implementation, program leaders and staff identified a need for care coordinators to learn

ways to help participants improve their motivation and change their behaviors to better manage their conditions. To address this need, program staff completed a course in motivational interviewing that provided guidance in how to engage participants and provide patient-centered care. Staff reported learning new skills to motivate their participants and help them reach their goals, and the low percentage of participants opting out of the program during the last year provides some evidence for their success in applying what they learned. AGH did not identify any other educational needs and did not conduct any additional training for the program.

d. Program time line

AGH hired new program staff, conducted training, and implemented the planned PCMH care coordination and care transitions program components according to the established time line. However, AGH experienced some early delays related to employing health IT to support the PCMH model. During the first year, development of the patient portal was interrupted when AGH changed to a new software vendor to accommodate required product functions. After AGH hired a new IT director, the portal development work progressed as planned. Of the 15 computers designated for installation at faith-based community organization partner sites to provide local access to the portal, 9 were installed in the second year of the program and 4 more were scheduled for installation during the first half of year three. Installation delays resulted from an initial lack of qualified support staff at partner sites to receive, install, and maintain the computers.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.4 summarizes the major facilitators and barriers to AGH's implementation effectiveness in each domain.

a. Program characteristics

Two characteristics of the AGH initiative helped the organization implement its program: (1) adaptability of the program to meet participants' and providers' needs and (2) providers' perceptions of the relative advantage of the program compared with the standard delivery of care. First, AGH continually adapted the PCMH program to reach more participants and better meet their needs. Early on, AGH expanded the target population to include patients with other conditions beyond the originally targeted diagnosis groups of COPD, CHF, and DM. This expansion has continued as providers identified patients with social needs and mental health conditions who could also benefit from care coordination. AGH added the KIT program to meet the needs of participants who no longer required care coordination services but could benefit from less intensive follow-up care to manage their health conditions. AGH also made process improvements throughout the duration of the program. For example, the care transitions team discovered that contacting patients during their hospital stays proved overwhelming to patients amidst the complex discharge process. In order to address this, the team modified enrollment procedures to include a brief in-hospital introduction followed by a program brochure mailed to the participant's home and a post-discharge telephone call. Finally, in order to improve outcomes

for participants transferred from AGH to a local skilled nursing facility, the care transitions coordinator developed a strong working relationship with the facility staff, participating in rounding for these participants and continuing follow-up for 30 days after discharge from AGH.

Table II.A.4. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Adaptability of the program to meet participants' and providers' needs • Providers' perceptions of the relative advantage of the program compared with the standard delivery of care • Frontline staff flexibility in applying the model to meet individual participants' needs	• Frontline staff flexibility contributing to a lack of standardization in applying the model
Implementation process	• Availability of resources to support the model • Engagement of program partners • Monitoring progress to guide ongoing improvement • Engagement of staff and providers	• Providers' perceptions of implementation burden and limitations of program impact for some participants • Turnover in partner staff dedicated to the program • Time-consuming and labor-intensive data collection and reporting process to support program monitoring
Internal factors	• Team communication and collaboration • Organizational structural features	• Limited reporting capacity related to health IT infrastructure
External factors	• New provider payment models	• Participants with complex needs and resource constraints

Sources: Interviews from second site visit. April 2015, and document review, June 2015.

Second, program staff and most providers recognized the relative advantage of the PCMH model in improving participants' care compared with the standard care delivery model. During the second site visit, one provider noted, "It gives the patients another layer of contact. They call their care coordinators without hesitation. The care coordinator can triage those calls and get patients the resources they need instead of the patient going to the ED and trying to see us when they may not need to. More than saving money, our patients are doing better." Another provider added, "I think I can do a better job with these [PCMH] resources. Now we can identify problems that were hidden to us before. It prevents a snowball effect of potential problems." However, it is important to note that some providers also shared concerns about long-term impacts of the program, as one provider interviewed during the second site visit expressed: "I have patients who improve when they're in the program. When they're better, they're discharged. But then they go back to their old ways. They know what to do, but they stop doing it after discharge. At some point they have to be accountable."

One feature of the AGH program, frontline staff flexibility in applying the model, acted as both a facilitator of and a barrier to implementation. AGH developed a checklist for care coordinators to follow for every participant, but permitted flexibility to adapt care plans and modify follow-up. Program staff emphasized the importance of this flexibility in enabling care

transition and care coordination nurses to exercise clinical judgment to customize participants' support. However, program administrators recognized a need to standardize care coordination services to build trust with providers and ensure that every participant received high quality care. During the second site visit, one provider stated, "I get different results with different care coordinators. It depends on the nurse and their comfort level and interest." Adding to the challenge, providers had different preferences for how much input and independent decision making they expected from a care coordinator, ranging from requesting notification for every participant complaint to wanting well-trained nurses to effectively triage participants' issues. Program administrators had planned to develop detailed participant protocols to help standardize services. But, in light of the large variation in participants' conditions and support needs, they moved away from this prescriptive approach and instead chose to pursue a more flexible disease-specific approach. AGH joined an accountable care organization (ACO) in January 2015, which gave them access to condition-specific evidence-based clinical pathways from a partner organization. AGH has reviewed these guidelines and aims to integrate them into its processes to improve the consistency of program services. However, it also recognized that employing clinical guidelines is not always straightforward, as one provider explained during the second site visit: "Protocols create a lot of conflict because 'best practice' is different based on whose evidence you're citing."

b. Implementation process

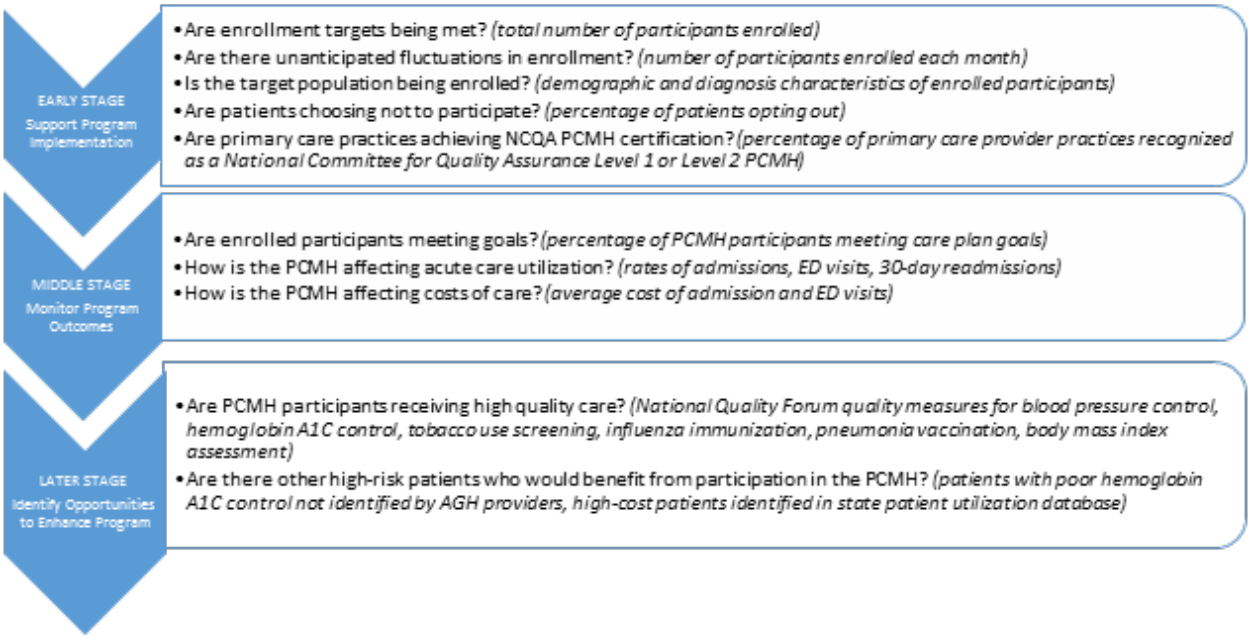
Three implementation process factors facilitated implementation of the AGH program: (1) availability of resources, (2) engagement of program partners, and (3) monitoring progress to guide ongoing improvement. First, program leaders used HCIA funding to increase capacity, which supported implementation of the PCMH program. Newly hired care coordinators provided patient-centered care coordination services that helped meet ongoing participants' needs between office visits while minimizing added burden on providers and practice staff. During the last year, the addition of a supervisor to oversee day-to-day program operations and a data specialist to manage data collection provided additional administrative support, enabling the program director to focus on process improvement and development of data-driven strategies to guide ongoing program enhancements.

Second, AGH established a strong collaboration with the WCHD to conduct participant home visits and assist them with social and financial needs. AGH encountered turnover of assigned program staff from the WCHD during the first year of the program, which caused temporary delays in delivering follow-up and home visits for care transitions participants, but new staff have remained in place since then. WCHD program staff are located in the same office as the AGH PCMH staff and have access to the AGH EHR, which has strengthened their communication and integration with the PCMH team. Staff, providers, and administrators agreed that the partnership with WCHD served as a critical component of the PCMH model, sometimes identifying issues impairing participants' ability to manage their conditions that they might not share with care coordinators or providers.

Third, AGH collected and monitored program metrics—including enrollment, utilization, and quality measures—throughout implementation of its PCMH model (Figure II.A.1). The process of collecting data and producing reports proved time-consuming and labor-intensive but

critical to informing program improvement decisions. In the early stage of implementation, AGH focused on process measures to monitor progress toward meeting enrollment targets and reaching its target population. When program leaders identified a dip in patient enrollment, they developed a report of patients with high utilization of acute care services to help providers identify patients who would benefit from participation in the PCMH. As the program progressed and enrollment increased, program leaders continued to monitor enrollment patterns but also began reviewing outcome trends. Investigation of an observed increase in readmissions revealed that most readmissions occurred among participants admitted from skilled nursing facilities. In response, AGH built a relationship with a local skilled nursing facility to provide care transitions support. Finally, AGH tracked a variety of quality measures to identify opportunities to improve the quality of participants' care. It successfully implemented an intervention to increase influenza vaccination rates after observing low vaccination rates.

Figure II.A.1. AGH HCIA program self-monitoring measures



Sources: Interviews from second site visit in April 2015 and document review June 2015.

One additional process factor, engagement of staff and providers, showed mixed influences on program implementation. Care coordination and care transitions staff consistently expressed their support for the PCMH model, recognized the program's potential positive impacts on participants, and valued the opportunity to build long-term relationships with participants. However, provider buy-in was more mixed. At the beginning of the program, providers voiced concerns about additional burden associated with implementing the model, but AGH alleviated these concerns by hiring new staff to deliver PCMH program services and to serve as the primary point of contact for participating patients to minimize burden on local practices. As the program progressed, administrators and care coordinators shared stories of positive impacts on individual participants, which helped keep providers engaged. However, several providers referred very few

patients to the program. Program staff and providers offered explanations for low referral volume, including a perceived lack of patient motivation to comply with program requirements; inconsistent quality of clinical support from the program team; and increased physician workload associated with nonreimbursed clinical management services—for example, guidance on medication adjustments and symptom management provided through care coordinators outside the office setting.

c. Internal factors

Characteristics of the organization implementing a program can influence implementation effectiveness. In the early stage of implementation, AGH program staff, providers, and administrators highlighted the importance of leadership commitment and their prior experience implementing a PCMH at two primary care practices in helping AGH implement its HCIA program. More recently, AGH staff emphasized the continued key roles of two other internal factors in facilitating program implementation: (1) team communication and collaboration and (2) organizational structural features.

First, program staff share a strong commitment to teamwork and have built a sense of camaraderie around the shared purpose of delivering high quality patient-centered care. Care coordinators and primary care providers communicate regularly through the EHR and during weekly in-person meetings to review participants' progress. Weekly team meetings provide opportunities for program staff to discuss day-to-day processes, share problems encountered, and coordinate schedules. In addition, in early 2015, AGH began hosting monthly grand rounds with program leaders, staff, and providers to provide a forum for review of participants with complex conditions and discussion of any program operational issues that arise. Program administrators reiterated the importance of providing these opportunities for staff and providers to actively collaborate and participate in decision making. Frontline staff expressed a high level of comfort voicing concerns to administrators and a belief that their perspectives mattered in guiding program improvement.

Second, administrators, care coordinators, and physicians believed that care delivery to participants through the PCMH model was facilitated by the relatively small size of the AGH organization (a 62-bed hospital and seven primary care practice sites) with most practices located in close proximity to one another and the program central office. Sharing offices in one location enabled AGH and WCHD program staff and administrators to meet informally, which facilitated timely problem solving and adjustments to staffing to accommodate work schedules and fluctuating caseloads. Physicians commented that face-to-face contact with care coordinators during weekly meetings and participant visits reduced the burden of reading and responding to high volumes of electronic communication and navigating complex care plans in the EHR.

One internal factor, technological infrastructure capacity, presented a barrier for AGH's program implementation. AGH has established inpatient and outpatient EHRs and built a patient portal, in part, to support its PCMH program. However, participants' data housed in five different databases made extracting the data and producing integrated reports to support program implementation difficult. During the second site visit, one staff member described the challenge they face: "The big barrier is data collection. We are a facility that needs help with EHRs and

databases speaking to each other and pulling high quality trustworthy data. It is very difficult to pull data. It is done by hand, and it is very time-consuming."

d. External factors

Features of an organization's external environment can also influence program implementation. Maryland is promoting adoption of new provider payment models, which has prompted strong leadership commitment to the PCMH model of care delivery and facilitated program implementation. Under a prior fee-for-service model, reduction of admissions and ED visits translated to financial losses for AGH. In January 2014, Maryland shifted to a global payment model that rewards hospitals for avoiding unnecessary hospitalizations. Administrators and executives believed the PCMH program would help them achieve savings under the new payment model, although no data are currently available to corroborate this belief. Participation in other shared savings programs, including an ACO that AGH joined in January 2015, offer additional opportunities to achieve savings to support implementation of the PCMH model.

Throughout implementation of the PCMH, AGH encountered challenges related to participants' needs and resources. Some participants in the target population faced significant barriers to care. Low-income participants, representing up to 50 percent of providers' caseloads on average, reportedly often had poor compliance with care plans, low literacy, financial constraints, limited access to transportation, and lack of caregiver support. Social work support provided by WCHD program staff helped connect participants to community resources to help meet these needs. However, AGH reported encountering additional barriers associated with providers' referrals of patients with mental health issues during the past year. Although they represented a small percentage of referrals, these patients reportedly often had complex medical and behavioral health needs and frequently did not respond to care coordination support or comply with care plans. After care coordinators made multiple attempts to engage these patients, many refused services and were discharged from the program. In addition, care coordinators did not have established relationships with local psychiatrists and other behavioral health providers, which limited their ability to connect participants with specialists to meet their mental health care needs.

4. Sustainability and scalability

From the beginning, AGH program administrators took a long-term view in adopting the PCMH model. AGH leaders believed the PCMH program offered an opportunity to build a model of care delivery that would enable them to reduce unnecessary use of acute care services and help them to benefit financially under Maryland's global payment model. Under the new payment model, adopted in January 2014 through an agreement with the Centers for Medicare & Medicaid Services (CMS), Maryland hospitals are rewarded for avoiding unnecessary hospitalizations. AGH anticipates that financial gains achieved under the revised payment structure will enable it to maintain support for the PCMH positions funded under the HCIA. Still, AGH recognizes the challenge it faces. As one program leader expressed during the second site visit, "Now keeping people out of the hospital is important for our bottom line. I feel confident that we're going in the right direction. The challenge is pacing with the payment structures. If we're doing the right thing but not getting paid for it, then we fail."

AGH's partnership with the WCHD ended when the program ended in June 2015. Staff, providers, and administrators agreed that the partnership with WCHD, which provided nursing and social work support to conduct home visits to participants, served as a critical component of the PCMH model. AGH administrators plan to pursue future collaborative agreements to continue to offer this service through a partnership rather than by hiring staff. As of June 2015, the organization continued to provide financial support for the program in its entirety, including funding it received through the Maryland Global Budget Revenue to support the program for the next two years. AGH also reported efforts to employ telemedicine to increase capacity of the PCMH team. AGH believes telemedicine will offer a new way to communicate with providers and participants that can increase face time with participants at home and during office visits and decrease time care coordinators spend on the road traveling to providers' offices. Said an administrator during the second site visit, "The most important thing is to try to educate our patients, especially high-risk patients. The medical home is trying to educate people to manage their conditions where they are—in their home, not in an institution."

AGH also continues to explore opportunities to expand the PCMH model as part of its population health management strategy. To reach more patients, including low-risk patients, it plans to engage more community organizations and build new relationships with independent outpatient providers. AGH estimated that most readmissions at AGH occur among patients treated by providers outside the AGH system. Therefore, program administrators are considering expanding enrollment in the care transitions program to all patients discharged from AGH, including those with non-AGH providers. "Some primary care providers don't know that their patients have been admitted at all, so it would benefit them and their patients. And we lower our readmission rates. It's a win-win," according to an administrator interviewed during the second site visit. Shared savings programs also offer opportunities to work with other providers. As noted earlier, AGH joined an ACO in January 2015 and, through this initiative, will begin working with more primary care providers and nurse practitioners to deliver coordinated care to participants. Finally, program leaders plan to use a statewide database to identify participants with AGH providers who are admitted to hospitals outside the AGH system. They aim to monitor these admissions to improve their ability to identify high-risk participants and track health care use by AGH participants.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected clinical sites or satellite offices provided important insights into the implementation process. Although these in-person interviews provided a rich source of data, views from the leadership and staff were limited to a small number of clinical locations and might not reflect the perspectives of clinicians practicing at other sites. In order to assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in fall 2014, the third year of the HCIA-funded program. Data from the survey provide additional insights into the implementation process and experience, as well as the contextual factors that might affect implementation effectiveness at AGH.

In this section, we report on AGH clinicians' views of their daily work life and practice. First, we focus on the contextual factors that can affect program implementation, including the characteristics of the practice location, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care team functions. We then present data on the alignment of AGH clinicians' views and experiences with the overall goals of the HCIA-funded program, as well as their awareness of and participation in the PCMH program and their view of the facilitators of and barriers to successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

Slightly more than half of the 13 AGH clinicians responded to the survey. Given the small sample size, we describe the results but do not present raw data. Respondents included a mix of physicians, nurses, and physician assistants, although not all respondents responded to every question. Clinical practice sites included group practices, a solo practice, and a community health center. Most AGH clinicians who responded reported that their primary source of compensation was either a fixed salary or a salary adjusted for performance.

AGH clinicians reported working in settings that are advanced in terms of health IT. Nationally, slightly more than half of physicians practice in settings with functional EHRs (Furukawa et al. 2014), but all AGH clinicians reported using health IT at their practice locations. All clinicians reported using electronic systems for ordering tests and procedures, accessing laboratory test results, prescribing medications, checking drug dosing and drug interaction alerts, and entering clinical notes. In addition, most responding clinicians use patient registries, a function that is not in widespread use nationally (DesRoches, Painter, and Jha 2014). Almost all AGH clinicians reported that their practices offer patient-facing technologies, providing their patients the option to do any of the following tasks online: request a prescription refill, request an appointment, or email a clinician about a medical question or concern.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall careers, levels of burnout, and perceptions of their practice environments can affect the success of program implementation and organizational change. All AGH clinicians responding to the survey were generally satisfied with their careers in medicine and did not report feeling burned out. They spent most of their time doing work that is well aligned with their training, with less than one-fourth of their time spent doing work that someone with less training could perform. The clinicians' ratings of their workplaces are summarized in Table II.B.1. Because fewer than 11 AGH clinicians responded to the survey, we did not report raw data.

Table II.B.1. Workplace ratings

The extent to which most clinician respondents agreed with each of the following statements	
Strongly agree	**Somewhat agree**
• Management is supportive of me. • I am encouraged to offer suggestions and improvements. • I have adequate opportunities to develop my professional skills. • The amount of work I'm expected to finish is reasonable. • Improving patients' capacity to manage their own care	• It is possible to provide high quality care to all of my patients

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Instead, this table shows the extent to which the majority of responding clinicians agreed with each statement.

AGH clinicians gave positive ratings to their workplace management. All respondents agreed that their management team is supportive, they feel encouraged by their supervisors to offer suggestions and improvements, they have adequate opportunities for professional development, and the amount of work they are expected to complete each week is reasonable.

The survey also assessed clinicians' beliefs about their ability to provide high quality care. All clinician respondents agreed with the statement, "It is possible to provide high quality care to all of my patients." The major barriers to providing optimal care reported by AGH clinicians included insufficient level of reimbursement, patients' difficulty with paying for care, and a lack of timely information about care provided to their patients by other physicians. Very few physicians reported that their ability to provide high quality care was hampered by a lack of research evidence to guide clinical decisions or by receiving excessive EHR communication. Results were mixed in three areas, with some physicians feeling somewhat limited : not having enough time to spend with patients during visits; having difficulty obtaining specialized diagnostic tests or treatments for patients in a timely manner; and having difficulty obtaining specialist referrals for patients in a timely manner.

c. Clinicians' perceptions of care team functioning

All AGH clinicians reported working as part of a care team. Respondents had positive perceptions of how care teams function. All clinicians responding either strongly or somewhat agreed that members of the care team relayed information in a timely manner, provided sufficient time for participants to ask questions during visits, and used common terminology when communicating with one another. Respondents were less supportive of the idea that team members verbally verified information they received from one another, and results were mixed about care teams following a standardized method of sharing information when handing off participants.

d. Alignment with goals of PCR

The survey asked clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The views of AGH clinicians generally aligned with the goals of PCR, with most clinicians rating 12 of the 13 goals as extremely important (Table II.B.2). All responding clinicians rated improving care continuity in primary care and improving patients' capacity to manage their own care as extremely important. However, only a few AGH clinicians rated increasing the use of EHRs and other health IT as an extremely important goal.

Table II.B.2. Importance of PCR goals

Proportion of clinicians who rated the following goals are extremely important:		
All	**Most**	**Few**
• Improving care continuity in primary care • Improving patients' capacity to manage their own care	• Improving care coordination for patients with chronic conditions • Improving appropriateness of care • Reducing ED visits • Reducing overall health care spending • Increasing access to primary care • Increasing the use of evidence-based practice in clinical care • Reducing hospital readmissions • Improving the capability of health care organizations to provide patient-centered care • Improving the capability of health care organizations to provide team-based care • Increasing the number of primary care practices functioning as a PCMH	• Increasing use of EHRs and other health IT

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the AGH program was to change the way care was provided. Program administrators believe that clinicians are critical to that process. Understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on participants' outcomes. For example, if clinicians are aware of the program, have received appropriate and effective training, and believe that the AGH program will have a positive effect on the care they provide, they are likely to feel more invested in the program's success. Alternatively, those who feel more negatively about the program might be less likely to enthusiastically implement the intervention. In this section, we report on clinicians' experiences with and perceptions of the AGH program.

a. Awareness of the program and receipt of training

All AGH clinicians who responded to the survey were familiar with the PCMH program, and most reported receiving training related to the program. On average, responding clinicians received 7.2 hours of program-related training.

b. Perceived effect of program on participants' care

Nearly all responding clinicians believed the AGH program would have a positive effect on the quality of care they provide. Most also felt the program would improve the safety, efficiency, and patient-centeredness of the care they provide, as well as their ability to respond to participants' needs in a timely way. No clinician perceived a negative impact of the program; rather, some believed the intervention would have no effect on the care they provide or that it was simply too soon to tell.

c. Barriers to and facilitators of program implementation

Finally, we asked clinicians who were at least somewhat familiar with the PCMH program to rate the effect of a series of facilitators and barriers to program implementation (Table II.B.3). Most respondents rated three areas as having a positive impact on program implementation: availability of personnel, availability of relevant participant information at the point of care, and availability of community resources to care for participants with complex conditions. Only a few respondents reported barriers to program implementation, citing the amount of required documentation and the amount of time required by the program as having a negative impact. Respondents had mixed opinions about the effect of the other factors asked about in the survey, such as funding, technology, availability of evidence-based clinical information, and communication with other providers.

4. Conclusions about clinicians' attitudes and behavior

Clinician respondents reported being familiar with the program, and they generally believed that the program will positively impact their ability to provide high quality patient-centered care to program participants. The goals of the program also align with the goals of clinician respondents, including improving participants' capacity to manage their own conditions and improving care continuity. These findings are consistent with what providers reported during the second site visit.

Table II.B.3. Barriers to and facilitators of program implementation

Facilitators	Barriers	Mixed
• Availability of personnel • Availability of relevant participant information at the point of care • Availability of community resources to care for participants with complex conditions	• Amount of required documentation • The amount of time required by the program	• Level of program funding • Required use of computer and communications technology • Availability of evidence-based clinical information • Quality of interpersonal communications with other providers • Quality of interpersonal communications with specialists • Quality of interpersonal communications with other allied health professionals

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

C. Impacts on patient outcomes

1. Introduction

In this part of the report, we draw preliminary conclusions based on available evidence about the impacts of the care transitions component of AGH's PCMH program on Medicare beneficiaries' outcomes in three domains: quality-of-care outcomes, service use, and spending. Although the care transitions component serves Medicaid beneficiaries and Medicare beneficiaries enrolled in managed care plans as well as Medicare fee-for-service (FFS) beneficiaries, due to limitations in available data we have analyzed outcomes only for the Medicare FFS population (including those who are dually eligible for Medicare and Medicaid). Results might not be generalizable to the full population that the care transitions component serves. We first describe the methods for estimating impacts (Section II.C.2) and then the characteristics of the treatment beneficiaries (Section II.C.3). We next demonstrate that the treatment beneficiaries were similar at the start of the intervention to the beneficiaries we selected as a comparison group, which is essential for limiting potential bias in impact estimates (Section II.C.4). Finally, in Section II.C.5, we describe the quantitative impact estimates, their agreement with implementation findings, and our conclusions about program impacts in each domain. Our conclusions in this report are preliminary because the analyses do not yet cover the full period that we will include in the final impact analysis in future reports.

In consultation with CMMI, we decided not to attempt to estimate the impacts of AGH's care coordination program, the other main component of the AGH PCMH intervention. The small number of practices participating in the care coordination program means that our statistical models could not reliably detect even very large impacts. Further, the process that AGH used to identify and enroll beneficiaries into the care coordination program cannot be fully replicated in Medicare claims, making it difficult to define a credible comparison group. We also decided not to evaluate the smaller Keeping in Touch component of the PCMH program for similar reasons.

2. Methods

a. Overview

We estimated program impacts using a difference-in-differences framework. To implement this framework, we defined two cohorts of Medicare beneficiaries: (1) a *post-intervention cohort,* which included beneficiaries discharged from the hospital after the program began on January 1, 2013, and who met the program eligibility criteria (the post-intervention treatment group) and their matched comparison beneficiaries (the post-intervention comparison group); and (2) a *pre-intervention cohort,* which included beneficiaries discharged at least six months before the intervention began but otherwise met the program eligibility criteria (the pre-intervention treatment group) and their matched comparison beneficiaries (the pre-intervention comparison group). To estimate the program's impact during each intervention quarter following the qualifying hospital discharge, we (1) calculated the difference in outcomes between the post-intervention treatment and comparison groups that quarter and (2) subtracted any difference in outcomes between the pre-intervention treatment and comparison groups in the corresponding quarter. For example, when estimating impacts in the first three months after discharge, we calculated the difference in outcomes during those three months for the post-intervention treatment and comparison group, and then subtracted any difference during the first three months after discharge for the pre-intervention treatment and comparison group. This difference-in-differences approach helps to isolate the program impacts from any differences in post-hospitalization outcomes that existed between the treatment and comparison groups before the intervention began.

To focus the impact analyses, we specified a limited number of primary tests before examining any impact results. Each primary test defined an outcome, population, time period, the direction of expected effects for which we hypothesized to see impacts if the program is effective, and thresholds that we count as substantively important. We provided the awardee and CMMI an opportunity to comment on the primary tests and revised them as appropriate. We drew preliminary conclusions about impacts in each domain based on the results of these primary tests and the consistency of the primary test results with the implementation findings, reported in Section II.A, and secondary quantitative tests (robustness and model checks).

b. Treatment group definition

Post-intervention treatment group. The post-intervention treatment group includes Medicare FFS beneficiaries who met two criteria. First, they had to meet AGH's program eligibility criteria, to the extent that we could replicate them in claims. That is, the beneficiary had to (1) be discharged from AGH from February 1, 2013 (the date AGH enrolled its first patient into the care transitions component of its program) and September 30, 2014 (to allow at least three months of follow-up to the end of the outcome period for this report, December 31, 2014); and (2) be an AGH patient. We identified AGH patients as those who had their most recent primary care visit with an AGH provider (we received the list of providers from AGH) or who had the plurality of their primary care visits in the past two years with an AGH provider. The second criterion is that a beneficiary had to be continuously enrolled in FFS Medicare for the four quarters before his or her qualifying discharge. This restriction improved the matching

of treatment to potential comparison beneficiaries by ensuring we could use a full year of claims to develop baseline indicators of services use and diagnoses for matching.

This claims-based treatment group definition has two advantages over an alternative definition that includes only those who actually enrolled in the care transitions component of AGH's program. First, because AGH targeted any patients discharged from AGH with an AGH primary care provider (PCP), this definition corresponds to everyone the program intended to treat (that is, the definition follows an intent-to-treat design). Second, we can use exactly the same definition to identify a pre-intervention treatment group, which is needed to implement the difference-in-differences design. Although the intent-to-treat results are most relevant for policymakers, some stakeholders could be interested in impacts among only those who received the treatment. When comparing our treatment group definition to the roster of actual AGH enrollees, we found that 63 percent of the treatment group members were actually enrolled in the program. Therefore, any impacts measured among the full treatment group might understate the impacts among only those who actually enrolled. We did not conduct sensitivity analyses to estimate impacts among only those who enrolled because, without the ability to replicate individuals' enrollment decisions using claims data, we could not create a comparison group that would have made such sensitivity analyses meaningful.

In addition to defining the treatment group, we defined an enrollment date for each treatment group beneficiary, recognizing that not all members of the treatment group actually enrolled in the program. We defined the enrollment date as the day after the hospital discharge that qualified a person for the treatment group. If a beneficiary had multiple qualifying discharges during the study period, we selected the earliest one. The enrollment date serves as the anchor for defining the intervention quarters (for example, the first intervention quarter [I1] is the three months after the enrollment date) and for defining a beneficiary's baseline characteristics (which are defined on the enrollment date, or over the 12 to 36 months before it).

Pre-intervention treatment group. We defined the pre-intervention group using the same claims-based rule as for the post-intervention group, with one difference. The beneficiary had to be discharged from AGH from July 1, 2011, to June 30, 2012, allowing each beneficiary to be followed up for at least six months before the intervention began.

Additional sample restrictions in each quarter. To be included in the analytic sample in any given quarter, each treatment group member had to meet three additional criteria to contribute an observation for the quarter. First, the end of the quarter had to be no later than December 31, 2014, allowing the beneficiary to be potentially followed up for the full quarter using the claims data that were available for this report. Second, the beneficiary's outcomes had to be observable in Medicare claims for at least one day during the quarter. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer (including beneficiaries who are dually eligible for Medicaid). Finally, all of a treatment beneficiary's matched comparison beneficiaries (see next section) also had to be in the sample during the quarter, so that the treatment beneficiary's outcomes could be compared with the outcomes for all of his or her comparison beneficiaries.

c. Comparison group definition

Post-intervention comparison group. We constructed a comparison group of Medicare FFS beneficiaries who were similar to the post-intervention treatment group beneficiaries. This section describes how we constructed the matched comparison group whereas Section II.C.4 shows the balance we achieved between the two groups on the matching variables. We constructed the comparison group through three steps:

First, we identified a pool of *potential* comparison members. This pool consisted of all Medicare FFS beneficiaries (1) discharged from February 1, 2013, to September 31, 2014, from Peninsula Regional Medical Center (PRMC) in Salisbury, Maryland, which is about 30 miles from AGH, but where the care transitions component was not implemented; or (2) discharged from AGH (in the same time frame) but not attributed to an AGH provider (so the beneficiaries were not assigned to the post-intervention treatment group). We set the day following hospital discharge as the potential comparison beneficiary's pseudo-enrollment date. If a potential comparison beneficiary was discharged more than once, we set his or her pseudo-enrollment date to the day after the first discharge.

Second, we used the Medicare Enrollment Database and a beneficiary's Medicare claims in the 12 to 36 months before his or her pseudo-enrollment date to develop baseline characteristics for each beneficiary.

Finally, we used propensity score matching and exact matching techniques to limit the potential comparison pool to a list of matched comparison beneficiaries. Matching aims to reduce selection bias in observational studies by selecting comparison beneficiaries from the pool who are roughly equivalent to the treatment group across key baseline characteristics. The goal of matching is to achieve baseline equivalence between the treatment and matched comparison groups on the variables included in the matching process (Stuart 2010). For AGH, we matched on demographic characteristics, Medicare and Medicaid dual enrollment, original reason for Medicare entitlement, health status and chronic conditions, service use and spending 3 months before enrollment or pseudo-enrollment, and service use and spending 4 to 12 months before enrollment or pseudo-enrollment. Because service use and spending before enrollment or pseudo-enrollment are important predictors of these outcomes in the post-intervention period, the consensus is to use the baseline outcomes for matching.

Within the family of propensity score matching methods, we implemented a technique called *full matching* to form matched sets that contain one treatment and one or more comparison beneficiaries. The important benefit of full matching is that it achieves maximum bias reduction on observed matching variables and, subject to this constraint, maximizes the size of the comparison sample (Rosenbaum 1991; Hansen 2004). Each treatment beneficiary was matched to up to five beneficiaries from the potential comparison group.

We used exact matching techniques to ensure matched comparison group beneficiaries had (1) a qualifying inpatient discharge within 90 days of the treatment beneficiary's enrollment date, (2) the same gender as the treatment beneficiary, and (3) the same reason for the hospitalization that caused a person to enter the treatment or comparison group. Specifically, we used 19 unique

modified diagnosis-related group (MDRG) codes to define the types of hospital stays for most treatment beneficiaries. For the remaining treatment group beneficiaries, MDRG codes were too uncommon to provide sufficient matches in the comparison group; in this case, major diagnostic category (MDC) codes (instead of MDRG codes) were used for exact matching.

Pre-intervention comparison group. We constructed a comparison group of Medicare beneficiaries who were similar to the pre-intervention treatment group beneficiaries. The pool of potential comparison members consisted of all Medicare FFS beneficiaries (1) discharged from PRMC from July 1, 2011, to June 30, 2012; or (2) discharged from AGH (in the same time frame), but not attributed to an AGH provider. Because the sample sizes were smaller in the pre-intervention period, we exactly matched on 15 MDRG codes (instead of 19). Otherwise, the methods for constructing the pre-intervention comparison group were the same as we described earlier for the post-intervention comparison group. Five pre-intervention treatment beneficiaries were dropped because they could not be matched to any potential comparison beneficiaries.

Additional sample restrictions in each quarter. To be included in the analytic sample, a comparison group beneficiary had to meet the same additional criteria as the treatment group members—that is the end of the quarter had to be no later than December 31 and the beneficiary had to be observable in Medicare claims for at least one day of the quarter. Further, the comparison beneficiary's matched treatment group beneficiary also had to be in the sample during the quarter, so that the comparison beneficiary's outcomes could be compared to the outcomes for his or her treatment beneficiary.

d. Construction of outcomes and covariates

We processed Medicare claims and enrollment data to develop two types of variables: (1) **outcomes,** defined for each person in each intervention quarter during which they are members of the treatment or comparison group; and (2) **covariates** that describe a beneficiary's characteristics at the time of enrollment or pseudo-enrollment and were used in the regression models for estimating impacts to adjust for existing characteristics. We used one set of baseline covariates, without updating them each quarter, to avoid controlling in each intervention quarter for previous quarters' program effects, as this would bias the effect estimates away from detecting true impacts. For the post-intervention cohort of beneficiaries, the Medicare claims covered services provided from four years before the start of the intervention (February 1, 2009) to the end of the outcome period for this report (December 31, 2014). For the pre-intervention cohort, the claims cover services from July 1, 2007, to December 31, 2012. We ended on December 31, 2012, to avoid including outcomes for the pre-intervention cohort that actually occurred during the intervention period. Appendix 1 provides details on the methods we used to construct the outcome variables.

Outcomes. We calculated four quarter-specific outcomes that we grouped into three domains:

1. Domain: Quality-of-care outcomes
    a. 30-day unplanned readmission rate (percentage); for each person in the sample, this is a binary variable that equals one if the beneficiary had an unplanned readmission within 30 days of the discharge that qualified him or her for the treatment or comparison group, and zero if not[1]
2. Domain: Service use
    b. All-cause inpatient admissions (number/quarter)
    c. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
    d. Total Medicare Part A and B spending ($/month)

CMMI has specified all four of these outcomes as core outcomes for the evaluations of all HCIA programs.

Covariates. The covariates, defined at the enrollment (treatment group) or pseudo-enrollment date (comparison group) include (1) demographics (age, age-by-gender interactions, race and ethnicity, lives in a zip code with a poverty rate of 20 percent or higher); (2) whether dually enrolled in Medicare and Medicaid; (3) original reason for Medicare entitlement (old age, disability, or end-stage renal disease); (4) the number of months with Part A and B coverage 4 to 12 months before a beneficiary's pseudo-enrollment date; (5) Hierarchical Condition Category (HCC) score, which is a continuous score that CMS developed to predict a beneficiary's future Medicare spending; (6) whether a beneficiary has each of six chronic conditions (cancer, congestive heart failure, chronic obstructive pulmonary disease, chronic kidney disease, or Alzheimer's disease-related disorders, or senile dementia), created by applying Chronic Condition Warehouse algorithms to claims in the 12 to 36 months (depending on the condition) before the beneficiary's enrollment or pseudo-enrollment date; and (7) service use and Medicare Part A and B spending in the prior 3 months, and 4 to 12 months. Service use includes the number of unplanned readmissions, the number of inpatient discharges, the number of ED visits, and an indicator for one or more primary care physician visits.

e. Regression model

We used a regression model to implement the difference-in-differences design for estimating impacts. For each quarter-specific outcome, the model estimates the relationship between the outcome and a series of predictor variables, assuming that each of the predictor variables has a linear (additive) relationship with the outcome. The predictor variables include the beneficiary-

[1] This outcome takes on the value of missing if the stay that qualified a person for the treatment or comparison group does not meet the criteria for an index stay in the unplanned readmission measure (see Appendix 1).

level covariates (defined in Section II.C.2.d); an interaction of each beneficiary-level covariate with each intervention quarter; indicators for each matched set (a treatment beneficiary plus his or her matched comparison beneficiaries) in each quarter; whether the beneficiary is assigned to the treatment or comparison group; an interaction of a beneficiary's treatment status with an indicator for being in the post-intervention period (as opposed to the pre-intervention period); an interaction of a beneficiary's treatment status with each intervention quarter; and an interaction of a beneficiary's treatment status with each intervention quarter interacted a second time with an indicator for being in the post-intervention period. Appendix 2 provides details on the regression methods, including descriptions of the weights each beneficiary receives in the model and how the regressions account for correlation in outcomes across quarters for a given individual, and across individuals in the same matched set.

The estimated relationship between the interaction term and an outcome in a given quarter gives the difference-in-differences estimate for that quarter and outcome. It measures the average difference between outcomes for post-intervention beneficiaries assigned to the treatment and comparison groups in a certain quarter, subtracting out any differences between the pre-intervention treatment and comparison groups during the same quarter. By providing separate difference-in-differences estimates for each quarter, the model enables the program's effects to change with the length of time the beneficiaries are enrolled in the program (which is expected to occur). We can also test impacts over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the difference-in-differences estimates, allowing for statistical tests that determine whether observed differences are likely due to chance.

f. Primary tests

Table II.C.1 shows the primary tests for AGH, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important (expressed as a percentage change from the counterfactual—that is, the outcome the treatment group would have had in the absence of the HCIA-funded intervention). The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for detail and a description of how we selected each test).

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** AGH's central goal is to reduce hospitalizations, ED visits, and Medicare Part A and B spending, so our primary tests address these three outcomes. In addition, the primary tests address one quality-of-care outcome the intervention is expected to affect: 30-day unplanned hospital readmissions. AGH's original HCIA proposal contained no separately stated goals for the care transitions component. Therefore, we assume that AGH's target outcomes for the care transitions program are the same as those for the HCIA program as a whole.
- **Time period.** AGH's proposal contained no specific time frame for reaching the program goals, but the literature on transitional care interventions indicates effects on readmissions

tend to be concentrated in the period following an initial, or index, hospital discharge (Peikes et al. 2012). For this reason, the primary tests measure impacts on the readmission rate in the 30 days following the (index) inpatient admission associated with the beneficiary's enrollment or pseudo-enrollment (that is, the stay that qualified a person for the treatment or comparison group).[2] Similarly, we expect effects for the other three outcomes—hospitalizations, ED visits, and spending—to be concentrated in the first one to three months following the enrollment admission. For these three outcomes, however, we set the time period for the primary tests to the first two quarters immediately following the enrollment admission, because some studies show impacts of transitional care programs over longer periods (Peikes et al. 2012).

- **Population.** AGH expected to have impacts for the population of beneficiaries enrolled in the care transitions component of its program. Therefore, the primary tests include all (observable) Medicare FFS beneficiaries who met the care transitions component's enrollment criteria. Although AGH did enroll patients with non-Medicare insurance, we do not have data to cover patients with commercial insurance or no insurance. We do not include Medicaid beneficiaries in our primary tests (unless they are also enrolled in Medicare) because we do not expect, by the end of the evaluation, to have Medicaid data that is timely enough to cover the primary test period for any (or a substantial number of) beneficiaries.
- **Direction (sign) of the impact estimate.** The primary tests are testing for a reduction, relative to the counterfactual, for each of the four outcome measures. We do this because the hypothesis is that the program will have favorable effects on the outcomes, which in all cases means reducing them relative to the counterfactual.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we have pre-specified thresholds for what we call substantive importance. The 11.6 to 15.0 percent thresholds we chose for substantive importance (depending on the outcome) are 75 percent of AGH's expected effects. (We use 75 percent recognizing that AGH could still be considered successful if it approached, but did not achieve, its fully anticipated effects.) The 15 percent threshold for the readmission rate is extrapolated from the literature (Peikes et al. 2012), because AGH did not specify by how much it expected to reduce these hospitalizations.

[2] The time period is defined this way because the matching variables were balanced for the treatment and comparison groups at the beneficiaries' enrollment or pseudo-enrollment dates due to our matching approach, but they would not have been balanced for subsequent inpatient admissions.

Table II.C.1. Specification of the primary tests for Atlantic General Hospital's care transitions component

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for pre-intervention differences)[b]	Population	Substantive threshold (impact as percentage of the counterfactual)[c,d]
Quality-of-care outcomes (1)	30-day unplanned hospital readmission rate (%)	The 30 days immediately following the enrollment admission[e]	All observable Medicare FFS beneficiaries attributed to the treatment group with a qualifying enrollment admission (index stay)[e]	-15.0%
Service use (2)	All-cause inpatient admissions (#/1,000/quarter)	Average over the first two quarters immediately following the enrollment admission[e]	All observable Medicare FFS beneficiaries attributed to the treatment group	-15.0%
	Outpatient ED visit rate (#/1,000/quarter)			-15.0%
Spending (1)	Medicare Part A and B spending ($/person/month)			-11.6%

Note: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

[a] We adjusted the *p*-values from the primary test results for the multiple comparisons made within each domain, but not across domains.

[b] The regression models for estimating program impacts controlled for differences in outcomes between the pre-intervention treatment and comparison groups.

[c] For all-cause hospitalizations, the outpatient ED visit rate, and Medicare spending, we set the substantive threshold to 75 percent of AGH's expected effect. For readmissions, an outcome for which AGH did not set an explicit target, we used Peikes et al.'s (2012) review of transitional care interventions as a guide when setting the threshold.

[d] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention.

[e] The enrollment admission is the inpatient discharge that led to a beneficiary being assigned to the treatment or comparison group.

AGH = Atlantic General Hospital; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

Due to limitations in data availability, we were able to conduct the primary tests in this report only partially. Specifically, we estimated impacts through December 31, 2014, for Medicare FFS beneficiaries enrolled before September 31, 2014. However the program did not end until June 30, 2015. Our third annual report will cover Medicare beneficiaries discharged from AGH through May 31, 2015, and include outcome data constructed with claims data through December 31, 2016. This definition will enable all treatment members to have potentially received at least one month of services before the program ended.

g. Secondary tests

We also conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the difference-in-differences design or random fluctuations in the data. We have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. Specifically, we repeated the primary tests above, but excluded from the sample 15 beneficiaries in the treatment group who were enrolled in the care coordination component of AGH's PCMH program, as well as their 58 matched comparison beneficiaries. If there were large differences between the primary tests and the secondary tests, it could suggest that impact estimates are being (fully or partially) driven by the care coordination component, not the care transitions component, of AGH's program.

h. Synthesizing evidence to draw conclusions

Within each domain, we drew one of four conclusions about program effectiveness, based on the primary test results, the results of secondary tests, and the plausibility of those findings given the implementation evidence. These four possible conclusions are as follows: (1) statistically significant favorable effect (the highest level of evidence), (2) substantively important favorable effect, (3) indeterminate effect, and (4) substantively important unfavorable effect. (We cannot conclude that a program has a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program harms.)

Our decision rules for each of the four possible conclusions are described in Appendix 3. In short, we concluded that a program had a statistically significant favorable effect in a domain if (1) at least one primary test result in the domain was favorable and statistically significant, after adjusting the statistical tests to account for multiple tests (if applicable) within a domain; or (2) the average impact estimate across all primary tests in the domain was favorable and statistically significant. In both cases, we had to also determine that the primary test results were plausible given the secondary tests and implementation evidence. We concluded that a program had a substantively important favorable effect if the average impact estimate was substantively important but not statistically significant, and if the result was plausible given the secondary tests and implementation evidence. In contrast, if the average impact estimate was unfavorable (opposite the hypothesized direction), larger than the substantive threshold, and unfavorable effects were plausible given the other evidence, we concluded the program had a substantively

important unfavorable effect. Finally, if the tests in a domain did not meet any of these criteria, we concluded that the impact in that domain was indeterminate.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the 460 beneficiaries in the post-intervention treatment group, which can be seen in the first column of Table II.C.2, Panel A, and the characteristics of the 226 beneficiaries in pre-intervention treatment group (before the program began), which can be seen in the first column of Table II.C.2, Panel B. For benchmarking purposes, the last column shows the values of relevant variables for the national Medicare population, when available.

Post-intervention treatment group. Some demographic characteristics of the 460 Medicare FFS beneficiaries in the post-intervention treatment group (such as gender and age) are similar to benchmarks for the national Medicare population, but other characteristics in Table II.C.2., Panel A indicate the treatment group has more health care needs than the general population. The HCC risk score for the treatment group is 2.61, indicating that the group can be expected to have Medicare spending that is 2.61 times higher than the national average (1.00) over the next year. The incidence of congestive heart failure, chronic obstructive pulmonary disease, and chronic kidney disease in the treatment group was more than twice the national average.

Treatment group members also had high service use (inpatient admissions and outpatient ED visits) and spending relative to national Medicare averages. For example, the treatment group beneficiaries had on average 1,109 hospitalizations (per 1,000 beneficiaries in the quarter) before their enrollment dates and 86 hospitalizations (per 1,000 beneficiaries per quarter) in the period 4 to 12 months before their enrollment dates, compared with a national average of 74. The program targeting criteria explain the spike in this utilization outcome in the quarter before pseudo-enrollment. The program enrolls people who are in the hospital; therefore, the population hospitalization rate must exceed 1,000 (corresponding to at least one stay per person) in that quarter. These hospitalizations, and perhaps other utilization, drove up Medicare spending as well.

Pre-intervention treatment group. Although the pre-intervention treatment group was not required to be the same as the post-intervention treatment group by construction, the two groups were largely similar to each other. The characteristics in Panel B of Table II.C.2 demonstrate the pre-intervention treatment group had significant health care needs, with average HCC scores of 2.73 and incidence of congestive heart failure, chronic obstructive pulmonary disease, and chronic kidney disease higher than the national average. The average service use and spending patterns over 12 months before enrollment for the pre-intervention treatment group were similar to patterns of the post-intervention treatment group. However, there were some differences between the two groups in the reasons for hospitalization.

Table II.C.2. Characteristics at baseline of treatment and comparison beneficiaries in the pre- and post-intervention cohorts for Atlantic General Hospital's care transitions component

Characteristic	Treatment group (n = 460)	Unmatched comparison pool (n = 7,334)	Comparison group (n = 1,976)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS average
Panel A: Post-intervention cohort						
Exact match variables[c]						
Male (%)	44.1	44.4	44.1	0	0	44.7[d]
Number of days from January 1, 2013, to enrollment	312.1	303.8	311.4	0.7	0.004	n.a.
Reason for hospitalization[e]						
MDRG 114: Intracranial hemorrhage or cerebral infarction (%)	4.1	5.4	4.1	0	0	NA
MDRG 409: COPD	4.6	3.7	4.6	0	0	
MDRG 410: Simple pneumonia and pleurisy (%)	7.6	6.3	7.6	0	0	NA
MDRG 524: Heart failure and shock (%)	5.4	5.8	5.4	0	0	NA
MDRG 1110: Renal failure (%)	4.1	3.2	4.1	0	0	NA
MDRG 1113: Kidney and urinary tract infections (%)	4.3	2.8	4.3	0	0	
MDRG 1808: Septicemia (%)	5.7	6.2	5.7	0	0	NA
Propensity matched variables[f]						
Demographic characteristics						
Age (years	76.68	74.79	76.42	0.26	0.024	71[g]
Race: white (%)	91.5	82.2	90.2	1.3	0.043	81.8[d]
Zip code poverty rate greater than 20 percent (%)	1.3	12.6	2.2	-0.9	-0.065	NA
Medicare-related characteristics						
Dual status at enrollment	13.0	22.1	14.0	-0.9	-0.026	22[h]
Original reason for entitlement (%)						
Disability	18.9	24.2	19.6	-0.7	-0.017	16.7[d]
ESRD	0	1.6	0.2	-0.2	-0.076	0.13[d]
Health status and chronic conditions						
HCC risk score	2.61	2.71	2.66	-0.05	-0.031	1.0
Chronic conditions[i] (%)						
Alzheimer's	9.3	6.7	7.4	1.9	0.073	4.9[j]
Cancer	16.7	17.5	18.1	-1.4	-0.037	NA
CHF	40.4	40.5	41.2	-0.7	-0.015	15.3[j]
COPD	31.3	33.2	32.6	-1.3	-0.028	11.8[j]
CKD	45.4	47.8	45.5	-0.1	-0.001	16.2[j]
Diabetes	43.7	44.1	44.8	-1.1	-0.022	28.0[j]
Alzheimer's disease, related disorders, or senile dementia	18.9	16.6	17.6	1.3	0.035	11.1[j]
Service use and spending 3 months before pseudo-enrollment						
Number of unplanned readmissions (#/1,000/quarter)	54	60	43	12	0.056	NA
Number of hospitalizations (#/1,000/quarter)	1,109	1,120	1,082	26*	0.092	0.074[k]
Number of ED visits (#/1,000/quarter)	398	380	388	9	0.011	0.105[l]
Primary care (%)[m]	96.1	95.4	95.6	0.5	0.026	NA
Medicare spending ($/month)	6,209	7,042	6,184	25	0.004	860[n]
Service use and spending 4 to 12 months before pseudo-enrollment						
Number of unplanned readmissions (#/1,000/quarter)	8	20	5	3*	0.081	NA
Number of hospitalizations (#/1,000/quarter)	86	119	82	3	0.018	0.074[k]
Number of ED visits (#/1,000/quarter)	245	254	241	5	0.009	0.105[l]
Primary care (%)[m]	95.0	85.7	94.3	0.7	0.030	NA
Medicare spending ($/month)	1,406	1,519	1,246	160	0.072	860[n]
Omnibus test for balance on matching variables[o]						
p-value	0.70					n.a.

Table II.C.2 *(continued)*

Characteristic	Treatment group (n = 226)	Unmatched comparison pool (n = 4,395)	Comparison group (n = 1,008)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS average
Panel B: Pre-intervention cohort						
Exact match variables[c]						
Male (%)	43.8	42.2	43.8	0	0	44.7[d]
Number of days from January 1, 2013, to enrollment	-364.1	-380.3	-366.9	2.8	0.028	n.a.
Reason for hospitalization[e]						
MDRG 114: Intracranial hemorrhage or cerebral infarction (%)	4.4	5.1	4.4	0	0	NA
MDRG 409: COPD	4.4	4.5	4.4	0	0	
MDRG 410: Simple pneumonia and pleurisy (%)	6.2	5.6	6.2	0	0	NA
MDRG 524: Heart failure and shock (%)	8.0	6.9	8.0	0	0	NA
MDRG 1110: Renal failure (%)	6.6	3.4	6.6	0	0	NA
MDRG 615: GI hemorrhage (%)	4.9	9.1	4.9	0	0	
MDRG 807: Major joint replacement (%)	4.4	5.1	4.4	0	0	NA
Propensity matched variables[f]						
Demographic characteristics						
Age (years)	77.95	75.67	77.52	0.42	0.040	71[g]
Race: white (%)	93.8	82.7	91.2	2.6	0.093	81.8[d]
Zip code poverty rate greater than 20 percent (%)	4.9	11.5	6.1	-1.2	-0.049	NA
Medicare-related characteristics						
Dual status at enrollment	9.3	20.3	11.5	-2.2	-0.070	22[h]
Original reason for entitlement (%)						
Disability	12.8	22.1	15.9	-3.1	-0.088	16.7[d]
ESRD	0	1.5	0.4	-0.4	-0.081	0.13[d]
Health status and chronic conditions						
HCC risk score	2.73	2.78	2.68	0.06	0.037	1.0
Chronic conditions[i] (%)					-0.011	
Alzheimer's	8.4	8.5	8.7	-0.3	0.041	4.9[j]
Cancer	22.6	17.7	20.9	1.7	0.035	NA
CHF	43.4	44.7	41.6	1.7	-0.009	15.3[j]
COPD	33.2	35.3	33.6	-0.4	0.033	11.8[j]
CKD	52.7	50.4	51	1.7	-0.030	16.2[j]
Diabetes	44.7	43.7	46.2	-1.5	0.021	28.0[j]
Alzheimer's disease, related disorders, or senile dementia	24.3	19.9	23.5	0.9	-0.011	11.1[j]
Service use and spending 3 months before pseudo-enrollment						
Number of unplanned readmissions (#/1,000/quarter)	27	67	24	3	0.017	NA
Number of hospitalizations (#/1,000/quarter)	1,084	1,129	1,078	6	0.022	0.074[k]
Number of ED visits (#/1,000/quarter)	296	375	303	-7	-0.012	0.105[l]
Primary care (%)[m]	96.9	95.7	95.9	1.0	0.051	NA
Medicare spending ($/month)	6,603	7,203	6,116	486	0.081	860[n]
Service use and spending 4 to 12 months before pseudo-enrollment						
Number of unplanned readmissions (#/1,000/quarter)	7	33	6	1	0.031	NA
Number of hospitalizations (#/1,000/quarter)	106	160	103	4	0.019	0.074[k]
Number of ED visits (#/1,000/quarter)	252	223	204	48*	0.136	0.105[l]
Primary care (%)[m]	95.6	87.0	93.5	2.0	0.083	NA
Medicare spending ($/month)	1,306	1,680	1,266	40	0.016	860[in]
Omnibus test for balance on matching variables[o]						
p-value	0.85					n.a.

Table II.C.2 *(continued)*

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code poverty rate merged from the American Community Survey ZIP Code Characteristics.

Notes: Characteristics are measured at the date of the inpatient discharge from AGH or PRMC that led to a beneficiary being assigned to the treatment or comparison group (the beneficiary's enrollment or pseudo-enrollment date). The post-intervention cohort includes beneficiaries whose enrollment or pseudo-enrollment dates were from February 1, 2013, to September 30, 2014, and the pre-intervention cohort includes beneficiaries whose enrollment or pseudo-enrollment dates were from July 1, 2011, to June 30, 2012.The comparison group means are weighted based on the number of matched comparisons per treatment beneficiary. For example, if four comparison beneficiaries are matched to one treatment beneficiary, each of the four comparison beneficiaries has a matching weight of 0.25.

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the matched treatment and comparison groups, which is pooled across the matched treatment and comparison groups.

[b] The standardized difference is the difference in means between the treatment and comparison groups divided by the SD of the variable, which is pooled across the treatment and comparison groups.

[c] Variables on which we required treatment and comparison members to match exactly. For example, a treatment group beneficiary whose reason for hospital discharge was intracranial hemorrhage or cerebral infarction (MDRG 1114) could be matched only to a comparison beneficiary who had the same reason for discharge.

[d] Chronic Conditions Data Warehouse (2014a, Table A.1).

[e] The reason for the hospitalization that caused a person to enter the treatment or comparison group. We used MDRG codes to define the types of hospital stays. In addition to the hospitalization types listed in the table, we exactly matched on 12 other MDRGs, which captured the reason for discharge for most treatment beneficiaries. For the remaining treatment group beneficiaries, MDRG codes were too uncommon to provide sufficient matches in the comparison group; in such cases, MDC codes (instead of MDRG codes) were used for exact matching. To pay acute care inpatient FFS claims, Medicare assigns discharges to Medicare severity diagnosis related groups (MS–DRGs) which group patients with similar clinical problems expected to require similar amounts of hospital resources; MDRGs group one or more related DRG codes into larger categories. MDC codes, in turn, group one or more MDRG codes together into even larger categories.

[f] Variables on which we matched through a propensity score, which captures the relationship between beneficiaries' characteristics and their likelihood of being in the treatment group.

[g] Health Indicators Warehouse (2014a).

[h] Health Indicators Warehouse (2014c).

[i] The chronic condition flags are calculated using one to three years of claims before the enrollment or pseudo-enrollment date (depending on the condition), using the Chronic Conditions Data Warehouse definitions.

[j] Chronic Conditions Warehouse (2014b, Table B.2).

[k] Health Indicators Warehouse (2014b).

[l] Gerhardt et al. (2014).

[m] Percentage of beneficiaries with any expenditures for primary care services in the 3 months before enrollment (or 4 to 12 months before enrollment).

[n] Boards of Trustees (2013).

[o] Results from an overall chi-squared test indicate the likelihood of observing a set of differences on the matching variables that is as large as what was observed if the treatment and comparison beneficiaries in the matched sample are equivalent on all the matching characteristics indicated. For example, the values of $p = 0.70$ for the chi-squared test for the post-intervention cohort suggests that the two groups are well balanced, because we cannot reject the null hypothesis that they are the same.

* Significantly different from zero at the .10 level, two-tailed test. No differences were significantly different from zero at the .05 or .01 levels. (Note: The primary tests assume one-tailed tests, for the reasons explained in the text.)

AGH = Atlantic General Hospital; CHF = congestive heart failure; CKD = chronic kidney disease; CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disease; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; GI = gastrointestinal; HCC = Hierarchical Condition Category; MDC = major diagnostic category; MDRG = modified diagnosis-related group; MV = mechanical ventilation; PCP = primary care provider; PRMC = Peninsula Regional Medical Center; SD = standard deviation.

NA = not available.

n.a. = not applicable.

4. Equivalence of the treatment and comparison groups at the start of the intervention

Demonstrating that the treatment and comparison groups are similar at the start of the intervention is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying the difference-in-differences design—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment group not received the intervention.

Post-intervention equivalence. Panel A of Table II.C.2 shows that the treatment and comparison beneficiaries in the post-intervention period were similar at baseline (that is, before enrollment or pseudo-enrollment). By construction, there were no differences between the two groups on the exact matching variables—gender, date of discharge, and the reason for enrollment. There were some differences between the treatment group beneficiaries and matched comparison group beneficiaries on the variables we matched through propensity scores, but the standardized differences across the propensity score matching variables are all well below our target of 0.25 standardized differences, and even within 0.10 standardized differences (the 0.25 target is an industry standard; for example, see Institute of Education Sciences 2014). The omnibus test that the treatment and comparison beneficiaries are perfectly matched on all variables cannot be rejected ($p = 0.70$), further supporting that the treatment and comparison groups are similar at the start of the intervention.

The propensity matching technique improved or did not affect the balance for most variables, but worsened the balance for a few variables. This can be seen in Panel A of Table II.C.2, which shows the means for the full comparison pool and for the selected comparison group. Key to our approach was improving balance on the reason for hospitalization (by MDRG or MDC), and the approach successfully removed *all* imbalance on this characteristic. Matching also improved the balance for other variables, particularly when the variables were imbalanced before matching (such as zip code poverty rate; original reason for entitlement; and the number of unplanned readmissions, hospitalizations, and ED visits 4 to 12 months before enrollment) because those variables had relatively more predictive power in the propensity score model. The improvements in balance on some variables came at the expense of increasing the differences between the treatment and comparison beneficiaries on (1) the percentage with cancer, (2) the percentage with diabetes, (3) the number of hospitalizations in the three months before enrollment, and (4) Medicare spending 4 to 12 months before enrollment. However, as mentioned earlier, the imbalance for all four variables was less than 0.10 standard deviations after matching.

Pre-intervention equivalence. Panel B of Table II.C.2 shows that the treatment and comparison beneficiaries in the pre-intervention period were also similar at baseline (that is, at pseudo-enrollment). We were able to exactly match comparison beneficiaries on gender, date of pseudo-enrollment, and reason for hospitalization. There were some differences between the treatment group beneficiaries and matched comparison group beneficiaries on the variables we matched through propensity scores, but the standardized differences across the propensity score matching variables were all well below our target of 0.25 standardized differences, and even within 0.15 standardized differences. The omnibus test that the treatment and comparison beneficiaries are perfectly matched on all variables cannot be rejected ($p = 0.85$).

5. Intervention impacts

In this section, we first present sample sizes and mean outcomes, by cohort and quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates; however, the differences in mean outcomes are not impact estimates by themselves. Next, we present the results of the primary tests (which are regression-adjusted), by domain. Then, we present the secondary tests results and assess whether the primary test results are plausible given the secondary tests. Next, we assess whether primary test results are plausible given the implementation evidence. We end with preliminary conclusions about program impacts in each domain.

a. Sample sizes

Post-intervention cohort. In the first intervention quarter (I1), the treatment group includes 460 treatment group beneficiaries and 1,976 comparison group beneficiaries (see Table II.C.3). This is the same sample that we used in matching, as shown in Table II.C.2. The sample decreases to 189 treatment group beneficiaries and 770 comparison beneficiaries in the second intervention quarter (I2). This drop in sample occurs because (1) some beneficiaries did not enroll or pseudo-enroll early enough to be followed up for the full second quarter, (2) some treatment or comparison group members exited the sample due to death or becoming unobservable, and (3) if any member of a matched set dropped from the sample, we—per the sample definitions—dropped all remaining members of the matched set.

Pre-intervention cohort. The treatment group in I1 includes 226 beneficiaries and the comparison group includes 481 beneficiaries. This is smaller than the I1 sample for the post-intervention cohort largely because the intake period for qualifying discharges is shorter for the pre-intervention cohort (365 days from July 1, 2011, to June 30, 2012) than for the post-intervention cohort (606 days from February 1, 2013, to September 30, 2014). As with the post-intervention cohort, and for the same reasons, the sample size drops from I1 to I2 for both the treatment and comparison groups.

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

Quality-of-care outcomes. The 30-day unplanned readmission rate for the comparison group members was 11.7 percent in the pre-intervention cohort and 15.1 percent in the post intervention cohort (Table II.C.3). The readmission rate was moderately lower (by 0.8 to 2.8 percentage points) for the treatment group than the comparison group, in both the pre- and post-intervention cohorts.

Table II.C.3. Sample sizes and unadjusted mean outcomes, by quarter, for Medicare FFS beneficiaries in the treatment and comparison groups, Atlantic General Hospital's care transitions component

	Number of Medicare FFS beneficiaries [a]			30-day unplanned hospital readmission rate (%)[b]			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/beneficiary/month)		
Quarter	T	C (un-weighted)	C (weighted)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Pre-intervention cohort															
I1	226	1,008	226	10.9	11.7	-0.8 (-6.7%)	385.0	349.1	35.8 (10.3%)	378.8	366.6	12.2 (3.3%)	$5,662	$3,949	$1,712 (43.4%)
I2	116	481	116	n.a.	n.a.	n.a.	232.8	184.3	48.4 (26.3%)	362.1	258.3	103.7 (40.2%)	$2,586	$2,117	$469 (22.1%)
Post-intervention cohort															
I1	460	1,976	460	12.3	15.1	-2.8 (-18.8%)	328.3	361.7	-33.4 (-9.2%)	343.5	353.4	-9.9 (-2.8%)	$4,344	$4,574	$-230 (-5.0%)
I2	189	770	189	n.a.	n.a.	n.a.	127.0	204.2	-77.2 (-37.8%)	375.7	273.3	102.4 (37.5%)	$1,963	$2,491	$-529 (-21.2%)

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Note: The quarters are three-month periods after a beneficiary's enrollment date (treatment group) or pseudo-enrollment date (comparison group), that is I1 is the first three months after enrollment or pseudo-enrollment, and I2 is months four to six. The means are weighted: each treatment group beneficiary receives a weight of 1; each comparison beneficiary receives a weight equal to the reciprocal of the total number of comparison beneficiaries who match to the same treatment beneficiary. The post-intervention cohort includes beneficiaries whose enrollment or pseudo-enrollment dates were from February 1, 2013, to September 30, 2014, and the pre-intervention cohort includes beneficiaries whose enrollment or pseudo-enrollment dates were from July 1, 2011, to June 30, 2012.

[a] The sample sizes are smaller in I2 than I1 because (1) some beneficiaries did not enroll or pseudo-enroll early enough to be followed up for the full second quarter, (2) some treatment or comparison group members exited the sample due to death or becoming unobservable, and (3) if any member of a matched set dropped from the sample, we—per the sample definitions—dropped all remaining members of the matched set.

[b] The sample sizes are smaller for the readmission outcome than the other outcomes, because the sample is limited to beneficiaries whose qualifying hospital discharges met the criteria for an index stay for the 30-day readmission measure (see Appendix 1). The sample sizes are 147 and 622 for the treatment and (un-weighted) comparison groups for the pre-intervention cohort, respectively; and 318 and (un-weighted) 1,307 for the treatment and comparison groups for the post-intervention cohort, respectively.

C = comparison group; CMS = Centers for Medicare & Medicaid Services; Diff = difference; ED = emergency department; FFS = fee-for-service; T = treatment group.

n.a. = not applicable.

Service use. For both the pre- and post-intervention cohorts, the mean hospitalization rates and outpatient ED visit rates in I1 and I2 for the comparison group were relatively high (for example, roughly 350 per 1,000 beneficiaries per quarter in I1 for both admissions and ED visits), signaling that patients remain vulnerable to acute events in the six months after hospital discharge. The hospitalization rates for the treatment group were 10 to 26 percent *higher* than the comparison group in the pre-intervention cohorts, but 9 to 38 percent *lower* than the comparison group in the post-intervention cohort. In contrast, outpatient ED visits were similar for the treatment and comparison groups in I1 for both the pre- and post-intervention cohorts. The treatment group rate in I2 was much (40 percent) higher than the comparison group's rate in the pre-intervention cohort, but was similarly much higher (37 percent) in I2 for the post-intervention cohort.

Spending. Medicare spending for the comparison group was higher in I1 than in I2, both for the pre- and post-intervention cohorts ($3,949 per beneficiary per month and $4,574 in I1 compared with $2,117 and $2,491 in I2). Spending was 22 to 43 percent *higher* in the treatment group than the comparison group in the pre-intervention cohort, but 5 to 21 percent *lower* in the post-intervention cohort.

c. Results for primary tests, by domain

Overview. Primary test results in the service use and spending domains represent the average impact estimate across the first two quarters of the intervention. In the quality-of-care outcomes domain, primary test results represent the impact estimate during the first 30 days of the intervention. Primary tests in the service use and spending domains indicate statistically significant favorable effects (Table II.C.4). The results for the quality-of-care outcomes domain were neither statistically nor substantively different from the regression-adjusted outcomes for the comparison group. As described earlier, these results are preliminary because the analyses do not yet cover the full period that we will include in the final impact analysis in future reports.

Quality-of-care outcomes. The treatment group's 30-day unplanned readmission rate (following enrollment) was 12.3 percent, 1.3 percentage points lower than the estimate of the counterfactual implied by the difference-in-differences regression model. (The estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.) This was a 9.7 percent difference, smaller than the substantive threshold (of 15.0 percent) and not statistically significant ($p = 0.36$). The statistical power values in Table II.C.4 (columns 5 and 6) imply that this analysis had limited power to detect small impacts on the readmission rate. In fact, the analyses (using a one-tailed test and a $p < 0.10$ cutoff) only had 23.5 percent power to detect a 15.0 percent impact on the readmission rate (the substantive threshold), and 43.5 percent power to detect a 30 percent impact, when the desired value is 80.0 percent power.

Table II.C.4. Results of primary tests for Atlantic General Hospital's care transitions component

Primary test definition					Statistical power to detect an effect that is[b]		Results			
Domain (# of test in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual)[a]	Size of the substantive threshold	Twice the substantive threshold	Treatment group mean	Regression-adjusted difference between treatment group mean and the counterfactual (standard error)[a,c]	Percentage difference[d]	*p*-value[e]
Quality-of-care outcomes (1)	30-day unplanned readmission rate (%)	The 30 days immediately following enrollment	All observable Medicare FFS beneficiaries attributed to treatment group with a qualifying enrollment admission (index stay)	-15.0%	23.5%	43.5%	12.3%	-1.3 p.p. (3.6)	-9.7%	0.359
Service use (2)	All-cause inpatient admissions (#/1,000/quarter)	The first two quarters immediately following enrollment	All observable Medicare FFS beneficiaries attributed to treatment group	-15.0%	35.9%	71.1%	227.6	-82.2* (50.6)	-26.5%	0.098[f]
	Outpatient ED visits (#/1,000/quarter)			-15.0%	33.1%	65.7%	359.6	8.4 (62.4)	2.4%	0.506[f]
	Combined (%)			-15.0%	49.1%	89.1%	n.a.	n.a.	-12.1% [g]	0.156[h]
Spending (1)	Medicare Part A and B spending ($/beneficiary/month)			-11.6%	42.1%	81.1%	3,153	-1,443*** (493)	-31.4%	0.002

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Notes: The results for each outcome are based on a difference-in-differences regression model that included one or two intervention quarter observations per beneficiary, as described in the text. For each quarter, the model calculates the regression-adjusted difference between outcomes for post-intervention period beneficiaries assigned to the treatment and comparison groups that quarter, subtracting out any differences between the pre-intervention treatment and comparison groups during the same intervention quarter. For three outcomes, the impact estimates from the first and second intervention quarters were averaged to obtain an average impact estimate for the first two quarters. The quarters are 91- or 92-day increments after the date of a discharge from AGH or PRMC that led to a beneficiary being assigned to the treatment or comparison group (the beneficiary's pseudo-enrollment date). For example, if a treatment beneficiary was discharged from AGH on July 15, 2013, and subsequently enrolled in the program on July 16, 2013, his or her first intervention quarter is July 16 through October 15, 2013; his or her second intervention quarter is October 16, 2013, through January 15, 2014. The estimates were adjusted for any differences in beneficiary-level covariates (defined in Section II.C.2.d) in each intervention quarter, and for indicators for each matched set (a treatment beneficiary plus his or her matched comparison beneficiaries) for each quarter.

Table II.C.4 *(continued)*

The treatment and comparison groups are limited to beneficiaries who were enrolled in FFS Medicare for each of the four quarters before the enrollment or pseudo-enrollment date. Furthermore, in each intervention quarter, the sample consists of Medicare FFS beneficiaries who were (1) enrolled early enough to be potentially followed up for all 91 or 92 days in the quarter and (2) whose outcomes were observable for at least one day during the quarter. The sample includes those who were in the sample for at least one of the intervention quarters. Outcomes are observable if the beneficiary is alive, enrolled in Medicare Part A and B, not enrolled in a comprehensive managed care plan, and has Medicare as his or her primary payer of medical bills. Outcomes are constructed through December 31, 2015. The sample sizes will change in future quarterly reports as new patients enroll in the program and the potential exposure period for prior enrollees increases. In each regression model, comparison group beneficiaries are weighted based on the number of matched comparisons per treatment beneficiary. For example, if four comparison beneficiaries are matched to one treatment beneficiary, each of the four comparison beneficiaries has a weight of 0.25. If either the treatment group beneficiary or *any* of the matched comparison group members in a matched set are not observable in a quarter, any remaining beneficiaries in the matched set are removed from the sample in that quarter.

[a] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[b] Statistical power is the probability of concluding that the program had a statistically significant favorable effect when the true effect was of the specified size. The power calculation is based on actual standard errors from analysis. For example, in the first row, a 15.0 percent effect on the readmission rate (from the estimated counterfactual of 12.3 + 1.3 = 13.6 percent) would be a change of 2.0 percentage points. Given the standard error of 3.6 percent from the regression model, we would be able to detect a statistically significant result only 23.5 percent of the time if the impact was truly 2.0 percentage points, assuming a one-sided statistical test at the $p = 0.10$ significance level.

[d] Percentage difference is calculated as the regression-adjusted difference-in-differences estimate, divided by the estimate of the counterfactual.

[e] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test).

[f] We adjusted the *p*-values from the primary test results for the multiple (two) comparisons made within the service use domain.

[g] The standard error for the combined percentage difference for the outcomes in the service use domain was 11.9 percentage points.

[h] This *p*-value tests the null hypothesis that the difference-in-differences estimates across the two outcomes in the service use domain, each expressed as percentage change from the comparison group mean, is greater than or equal to zero (a one-sided test).

*/**/*** Significantly different from zero at the .10/.05/.01 levels, one-tailed test, respectively. No difference-in-differences estimates were significantly different from zero at the .05 level.

AGH = Atlantic General Hospital; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Awards; p.p. = percentage points; PRMC = Peninsula Regional Medical Center.

n.a. = not applicable.

Service use. The treatment group averaged 227.6 all-cause inpatient admissions per 1,000 beneficiaries per quarter over the first two quarters following the beneficiary's enrollment or pseudo-enrollment date, which was estimated to be 82.2 admissions fewer than the counterfactual. This favorable difference between the treatment group mean and the counterfactual was statistically significant ($p = 0.098$, after adjusting for multiple statistical tests in the domain), and larger than the substantive threshold (26.5 percent versus 15.0 percent). The large difference is due to the fact that the treatment group's hospitalization rate was lower than the comparison group's during the intervention period, but higher than the comparison group's in the pre-intervention period, leading to a large difference-in-differences estimate. In contrast to hospitalizations, the rate of outpatient ED visits (per 1,000 beneficiaries per quarter) was 359.6 in the post-intervention period, only 8.4 visits below the counterfactual. This difference was small (a 2.4 percent difference, well below the substantive important threshold of 15.0 percent) and not statistically significant ($p = 0.51$). The mean percentage difference across the two outcomes was -12.1 percent (the average of -26.5 percent for inpatient admissions and +2.4 percent for ED visits), and was neither statistically significant nor substantively important. Analyses of the two outcomes in the service use domain were poorly powered to detect differences the size of the substantive threshold (with power less than 50.0 percent for each outcome individually or combined across both outcomes); the analysis of admissions was marginally powered to detect a difference of the size actually observed.

Spending. Medicare Part A and B spending for the treatment group averaged $3,153 per beneficiary per month over the first two quarters following the beneficiary's enrollment date, which was estimated to be $1,443 lower than the counterfactual. This favorable difference was statistically significant ($p = 0.002$), and also larger than the substantive threshold (31.4 versus 11.6 percent). The large difference is due to the fact that the treatment group's spending was lower than the comparison group's during the intervention period, but higher in the pre-intervention period, leading to a large difference-in-differences estimate. The analyses were poorly powered (42.1 percent) to detect differences the size of the substantive threshold, but well powered (> 80.0 percent) to detect differences of the magnitude actually observed.

d. Results for secondary tests

The results for the secondary tests were similar to those for the primary tests (Table II.C.5). There were statistically significant differences for all-cause inpatient admissions and Medicare Part A and B spending, similar in magnitude to the primary tests. For the readmission rate and outpatient ED visit rate the differences were similar in magnitude (smaller than the substantively important threshold) and were not statistically significant. The primary test results were plausible given these secondary tests; the secondary tests suggest that the care coordination component of AGH's program did not play a major factor in the primary test results.

Table II.C.5. Results of secondary tests for Atlantic General Hospital's care transitions component

Secondary test definition				Results			
Domain	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between treatment group mean and the counterfactual (standard error)	Percentage difference	*p*-value
Quality-of-care outcomes	30-day unplanned readmission rate (%)	The 30 days immediately following pseudo-enrollment	All observable Medicare FFS beneficiaries attributed to the treatment group with a qualifying enrollment admission (index stay) who were not enrolled in the care coordination component of AGH's program	12.6%	-1.2 p.p. (3.6)	-8.9%	0.369
Service use	All-cause inpatient admissions (#/1,000/quarter)	The first two quarters immediately following pseudo-enrollment	All observable Medicare FFS beneficiaries attributed to the treatment group who were not enrolled in the care coordination component of AGH's program	226.5	-83.5* (51.0)	-26.9%	0.051
	Outpatient ED visits (#/1,000/quarter)			348.8	-4.6 (62.4)	-1.3%	0.471
Spending	Medicare Part A and B spending ($/beneficiary/month)			3,153	-1,440*** (500)	-31.3%	0.002

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Notes: The analyses in Table II.C.5 were conducted in the same way as the analyses in Table II.C.4, except excluding 15 beneficiaries in the treatment group who were enrolled in the care coordination component of AGH's program, and their 58 matched comparison beneficiaries.

*/**/*** Significantly different from zero at the .10/.05/.01 level, one-tailed test. The *p*-values from the secondary test results were not adjusted for multiple comparisons within each domain or across domains.

AGH = Atlantic General Hospital; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; p.p. = percentage points.

e. Consistency of quantitative estimates with implementation findings

The primary test results were also plausible given implementation findings. The implementation findings suggest the care transitions component of AGH's PCMH program was reasonably well implemented and active during the post-implementation period. That is, we found no evidence that AGH failed to implement this component of the program. In fact, AGH exceeded its original enrollment targets and staffing target, and made process improvements throughout the duration of the program (for example, changes to the enrollment procedures; see Section II.A.2).

f. Conclusions about program impacts, by domain

Based on all evidence currently available, we have drawn the following preliminary conclusions about program impacts in each domain (as summarized in Table II.C.6)

- **Quality-of-care outcomes.** The program's impact on this domain was indeterminate because the primary test result for the single outcome in this domain (unplanned readmissions) was neither statistically significant nor substantially large. This indeterminate effect has two possible interpretations. First, the program may not have an effect for the population and period covered in this report. Alternatively, the program may have had an effect—and possibly even one that exceeded the substantive threshold—but, due to the statistical uncertainty in the estimate, we were unable to detect it.

- **Service use.** The program had a statistically significant favorable effect on service use. The primary test for all-cause inpatient admissions was favorable and statistically significant (after adjusting for two tests in the domain); the secondary tests confirmed the plausibility of the primary tests; and implementation findings indicate it is plausible that the care transition component was implemented in a manner that could have affected service use.

- **Spending.** The program had a statistically significant favorable effect on spending. The primary test for Medicare Part A and B spending was favorable and statistically significant. Furthermore, the secondary tests and implementation findings confirmed the plausibility of the results from the primary tests.

As mentioned previously, these conclusions are preliminary because the analyses do not yet cover the full period that we will include in the final impact analysis in future reports.

Table II.C.6. Preliminary conclusions about the impacts of Atlantic General Hospital's care transitions component on patient outcomes, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion		
		Primary test result(s) that supported conclusion	Primary test result plausible given secondary tests?	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	Indeterminate effect	None of the individual tests in the domain were statistically significant or substantively important	Yes	Yes
Service use	Statistically significant favorable effect	Estimate for all-cause inpatient admissions was favorable and statistically significant (after adjusting for two tests in domain)	Yes	Yes
Spending	Statistically significant favorable effect	Estimate for Medicare Part A and B spending was favorable and statistically significant	Yes	Yes

Sources: Tables II.C.4 and II.C.5.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

AGH received HCIA funding to implement its PCMH to offer care coordination for patients diagnosed with chronic conditions and post-hospitalization care transitions support for patients discharged from AGH with any diagnosis. AGH also developed an online patient portal to enhance communication between participants and providers and conducted education and outreach to support the PCMH model. The program aimed to reduce hospital admissions, ED visits, and total spending by helping participants manage their conditions and overcome social and financial barriers to care. Although AGH experienced some early delays related to employing health IT to support the PCMH model, AGH successfully implemented all program components, exceeding program enrollment and staffing targets and achieving high rates of participation and staff engagement. Key factors facilitating implementation included expanding the target population, staff perceptions of the advantages of the program for participants, flexibility to tailor support to an individual participant's needs, availability of resources to support the program operations, monitoring to identify process improvements, and Maryland's global payment model incentivizing the hospital to improve quality of care and reduce costs. AGH also worked to overcome several barriers, including a lack of existing data collection and reporting infrastructure, partner staff turnover, needs of participants with complex conditions and their noncompliance, and initial providers' perceptions of implementation burden. Clinician survey respondents echoed the opinions expressed by program administrators and care coordination and care transitions staff that the program will positively affect participants. Clinicians' views that the most important goals of PCR are to improve participants' capacity to

manage their own conditions and to improve care continuity aligned well with AGH's program goals to reduce overutilization of inpatient and ED services.

The impact evaluation found favorable and statistically significant impacts of the care transitions component on service use and spending for Medicare FFS beneficiaries during the first six months after beneficiaries were enrolled, but not on quality-of-care outcomes. The impact on service use was driven by a large impact estimate for hospitalizations that exceeded AGH's target of 20 percent, although there was no measurable effect for outpatient ED visits (the other outcome in the domain). The impact on spending exceeded AGH's target of 15 percent. However, we found no measurable effects for the one outcome in the quality-of-care domain (30-day readmission rates), which might be because the program had no effects on this outcome or because the statistical power to detect it was poor.

The next steps for this evaluation include (1) evaluating clinicians' attitudes and experiences with the program in the third year of the award through administered surveys; (2) extending the impact evaluation to include the entire period of the AGH program, which ended on schedule in June 2015; and (3) interpreting impact analysis findings in light of AGH's implementation experience.

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table A.1. Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

Chronic Conditions Data Warehouse. "Table B.2. Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014b. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf. Accessed November 19, 2014.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no. 9, 2014, pp. 1672–1679.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Hansen, Ben B. "Full Matching in an Observational Study of Coaching for the SAT." *Journal of the American Statistical Association,* vol. 99, no. 467, 2004, pp. 609–618.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Health Indicators Warehouse. "Medicare Beneficiaries Eligible for Medicaid (percent)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014c. Available at http://www.healthindicators.gov/Indicators/Medicare-beneficiaries-eligible-for-Medicaid-percent_317/Profile/ClassicData. Accessed August 4, 2015.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Peikes, Deborah, Rebecca S. Lester, Boyd Gilman, and Randall Brown. "The Effects of Transitional Care Models on Re-Admissions: A Review of the Current Evidence." *Generations*, vol. 36, no. 4, winter 2012–2013, pp. 44–55.

Rosenbaum, Paul R. "A Characterization of Optimal Designs for Observational Studies." *Journal of the Royal Statistical Society, Series B,* 1991, pp. 597–610.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for CareFirst BlueCross BlueShield

March 2016

Kristin Geonnotti
Greg Peterson
Lauren Hula
Boyd Gilman
Catherine DesRoches
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

CAREFIRST BLUECROSS BLUESHIELD

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by CareFirst BlueCross BlueShield (CareFirst) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the CareFirst program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on participants' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF CAREFIRST

CareFirst was originally awarded $20 million in HCIA funding to expand its commercial patient-centered medical home (PCMH) program to Medicare fee-for-service (FFS) beneficiaries in Maryland (Table I.1). The HCIA-funded initiative united the two largest payers in the region (CareFirst and Medicare) into a single health care financing model seeking to incentivize primary care providers (PCPs) to reduce health care costs while improving quality. The program's initial goal was to reduce hospital costs for Medicare FFS beneficiaries by 7.5 percent and total health care costs by 6.0 percent by the end of the award. However, because of a year-long delay in implementation, CareFirst now expects smaller effects in the third year of the award. CareFirst received a no-cost extension to continue providing program services until December 2015, with additional time through June 2016 to make incentive payments to panels.

Table I.1. Summary of CareFirst PCR program

Awardee's name	CareFirst
Award amount	$20,000,000[a]
Implementation date	August 1, 2013
Award end date	June 2016 (direct program services through December 2015)
Program description	CareFirst received HCIA funding to extend its commercial PCMH program to Medicare FFS beneficiaries in Maryland. Key components of the program include the following: 1. Develop and implement care plans for high-risk beneficiaries with multiple chronic conditions 2. Support these efforts with weekly care coordination and additional services, such as home-based assessments and in-home monitoring (as needed) 3. Use beneficiary-specific and population data to influence PCPs' behavior 4. Financially reward panels that reduce costs while improving quality
Innovation components	Care coordination, care transitions, payment incentives
Intervention focus	Panels
Workforce development	Created 44 new nurse local care coordinator positions, 4 nurse case manager positions, and 5 program consultant positions
Target population	Medicare FFS beneficiaries with chronic condition(s) who are high utilizers of health care services
Program setting	Provider-based (primary care practices)
Market area	Statewide (Maryland)

Table I.1 *(continued)*

Market location	Suburban, urban
Core outcomes (initial targets)	• 7.5 percent reduction in hospital costs for program participants • 6.0 percent reduction in total health care costs for program participants

Source: Review of CareFirst program reports, March 2015.

Note: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, recruiting panels, and undertaking other operational activities. Core outcome measures are based on CareFirst's initial targets.

[a] CareFirst was originally awarded $20 million to expand its PCMH program to Medicare beneficiaries in Maryland. An additional $4 million was allocated for use if CareFirst could find a partner to expand the program outside of Maryland, which did not happen.

For the HCIA-funded initiative, CareFirst recruited 52 practices that—for the purposes of CareFirst's commercial PCMH program—formed 14 medical care panels. Panels are groups of five to 15 PCPs (either physicians or nurse practitioners) who voluntarily agree to participate as a unit in terms of quality measurement and shared incentive payments. Panels can be formed by solo or small, independent group practices that agree to work together (referred to as a virtual panel); independent group practices that already fall within the size range; or a subsection of a large group practice. Health system-based practices, under common ownership of a hospital or health system, may also participate in the program. Only PCPs in the traditional primary care categories of internal medicine, family or general practice, geriatrics, and pediatrics can form panels.

II. SUMMARY OF FINDINGS

In this chapter, we summarize the methodology and present the main findings of the evaluation as they relate to (1) program implementation, (2) clinicians' attitudes and behaviors, and (3) participants' outcomes.

A. Program implementation

In this section, we first provide a detailed description of the program, highlighting how it has been adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of CareFirst's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in April 2014 and April 2015. We did not verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

CareFirst selected 14 panels that were already participating in its existing PCMH program for commercial patients, using HCIA funds to extend the model of care to FFS Medicare beneficiaries attributed to these panels. CareFirst's goal is to improve the quality and efficiency of care provided to Medicare beneficiaries in Maryland by transforming the way PCPs, including physicians and nurse practitioners, provide care to their participants. To help panels achieve these goals, the CareFirst's HCIA-funded initiative primarily focuses on care coordination for high-risk Medicare beneficiaries. The program hires nurse local care coordinators (LCCs) and places them in the primary care setting to help panels develop and implement care plans. It primarily includes weekly phone calls with participants to coordinate their care. Depending on the preferences of the PCPs, LCCs may also attend office visits with participants. Based on need, participants are also eligible to receive home-based health assessments, remote monitoring of health condition(s) at home, and behavioral health services (Table II.A.1).

The program design also includes two support elements that target the entire attributed population (as defined in Section II.A.1.b). First, program consultants hired by CareFirst help panels' PCPs throughout the program interpret population data by identifying utilization patterns, gaps in clinical care, and cost-savings opportunities. Second, CareFirst added a care transitions component in March 2014. Case managers reach out to participants to ensure post-acute and transitions of care needs are addressed.

Finally, CareFirst created financial incentives to foster program implementation and to hold panels accountable for the quality and cost outcomes of their attributed beneficiaries. PCPs receive $200 for developing a new care plan and $100 for updating an existing care plan. Medical panels that are able to keep the total cost of care for their attributed Medicare beneficiaries (not just participants with a care plan) below a specified target can earn an outcome incentive award (OIA), with the size of these incentive payments scaled to a panel's performance on quality measures.

b. Target population and patient identification, recruitment, and enrollment

CareFirst's care coordination program includes Medicare FFS beneficiaries with Parts A and B coverage with Medicare as their primary payer, and excludes patients who are dually eligible for Medicare and Medicaid. To be eligible for the program, beneficiaries must be attributed to a panel participating in the intervention. CareFirst conducts monthly attribution using claims data from CMS. A beneficiary is attributed to a panel if the practice's PCPs provided the plurality of a beneficiary's primary care services in the past 12 months (or in the past 24 months if the beneficiary received no primary care services in the past 12 months). CareFirst risk stratifies attributed beneficiaries to provide care coordination services to only those with the highest illness burden scores. Care transitions support can be provided to any attributed beneficiary who is not in a care plan and has had an acute care episode.

c. Service delivery protocols

The intervention protocols for care coordination and care transitions are described in Table II.A.1. In addition to these components, CareFirst also uses program consultants to help panels use data to focus on improving population health. Program consultants meet quarterly with each panel, but are often in more frequent communication with PCPs. For example, from January to March 2015, program consultants held 134 meetings with panels and PCPs.

Table II.A.1. Key details about program design and adaptation

	Program component	
	Primary: Care coordination	**Secondary: Care transitions**
Target population	CareFirst's care coordination program includes Medicare FFS beneficiaries with Parts A and B coverage with Medicare as their primary payer, and excludes patients who are dually eligible for Medicare and Medicaid. To be eligible for the program, beneficiaries must be attributed to a panel participating in the program. CareFirst conducts monthly attribution. A beneficiary is attributed to a panel if the practice's PCPs provided the plurality of a beneficiary's primary care services in the past 12 months (or in the past 24 months if the beneficiary received no primary care services in the past 12 months).	
Identification strategy	Care coordination services focus on attributed high-risk Medicare beneficiaries with multiple chronic and/or unstable conditions who are at the highest risk of hospitalization or other costly acute care services. CareFirst stratifies attributed beneficiaries into five illness bands based on their health status, using inpatient and outpatient diagnoses and demographic information to assess risk.[a] These illness bands are meant to help program staff identify beneficiaries who are most in need of additional care coordination services (and thus enrollment into care plans). LCCs and PCPs identify beneficiaries for care plans using illness burden scores, as well as through their judgment about who would benefit most from a care plan (such as PCPs' relationships with beneficiaries and their understanding of participants' medical and social needs).	Care transition services are targeted to all attributed beneficiaries who have been recently been admitted to the hospital. Services are provided to a smaller subset of beneficiaries who care managers reach and who agree to participate.
Recruitment/ enrollment strategy	LCCs and PCPs reach out to eligible participants primarily by telephone or during in-person visits to invite them to participate in the program. Medicare beneficiaries must verbally consent to program enrollment.	Case managers reach out to beneficiaries, primarily by telephone, to ensure all acute care needs are addressed.
Service delivery protocol	Patient-specific care plans describe a clinical strategy for each participant, typically implemented over the course of several months. LCCs are required to connect with beneficiaries in active care plans at least once per week (almost always via telephone), and are required to make at least three attempts to contact the participant. Direct participant care coordination via telephone can last 5 to 30 minutes, on average, depending on a participant's needs. PCPs review each care plan and its progress, depending on the participant's chronic condition or the timing of the participant's follow-up appointment with the PCP (which the LCC often attends).	Case managers assist beneficiaries in obtaining valuable resources following their hospitalization and, in several instances, have successfully transitioned the participant to an LCC for a longer-duration care plan. These activities are designed to avoid unnecessary re-hospitalizations and further breakdowns in their care. The case manager position is a relatively new addition to the HCIA-funded initiative; therefore, case managers do not follow well-defined intervention protocols at this time.
Service delivery protocol *(continued)*	Participants with care plans receive medication reconciliation and may receive several additional support services that CareFirst has phased in throughout the award. Based on need, participants are eligible to receive home-based health assessments, remote monitoring of health condition(s) at home, and behavioral health services. Medicare pays for any support services already reimbursed by traditional FFS Medicare (for example, home health), but CareFirst provides additional funding for the agency to import the data into CareFirst's portal. This provides PCPs and LCCs access to additional in-depth information about their participants.	

Table II.A.1 *(continued)*

	Program component	
	Primary: Care coordination	Secondary: Care transitions
Adaptations	None	Yes. Although not an initial part of the design, CareFirst began recruiting case managers in March 2014. The position was approved as part of the Year 1 carry-over funding request.

Source: Interviews from second site visit, April 2015; document review, March 2015.

[a] More specifically, CareFirst's illness burden score is based on its own diagnostic cost grouper classification model, which is similar to hierarchical condition categories scores.

d. Intervention staff and workforce development

Table II.A.2 provides key details about staff hired for the HCIA-funded initiative. Through its vendor Healthways, CareFirst hired registered nurse LCCs to facilitate the development and implementation of care plans for high-risk participants as defined by illness burden score. CareFirst added case managers in March 2014 to support the program by focusing on care transitions for a small subset of participants who were recently discharged from the hospital. CareFirst also hired program consultants at the start of the program, who analyze each panel's attributed participant population data.

Table II.A.2. Key details about intervention staff

Program component	Staff members	Staff /team responsibilities	Adaptations?
Care coordination	Local care coordinator	Through its vendor Healthways, CareFirst hired registered nurse LCCs to facilitate the development and implementation of care plans for high-risk participants. LCCs are supposed to contact participants in an active care plan at least once a week and are required to make at least three contact attempts each week. A full caseload for an LCC is considered to be 45 active care plans.	No
Care transitions	Case manager	Case managers, who are registered nurses, reach out to participants experiencing a care transition after an acute care episode to ensure all post-acute care needs are addressed. For example, case managers assist participants in obtaining resources available in the community. They also transition eligible participants to an LCC for a longer-duration care plan.	Yes. Although not a part of initial program implementation, CareFirst began recruiting case managers in March 2014. The position was approved as part of the Year 1 carry-over funding request.
Population health	Program consultant	CareFirst hired program consultants, who inform PCP behavior by providing them with provider- and panel-level data reports to identify key cost drivers, quality metrics, and potential gaps in care. Program consultants tend to focus on a panel's entire attributed population.	No

Source: Interviews from second site visit, April 2015; document review, March 2015.

2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators and self-reported information included in CareFirst's quarterly self-monitoring and measurement reports.

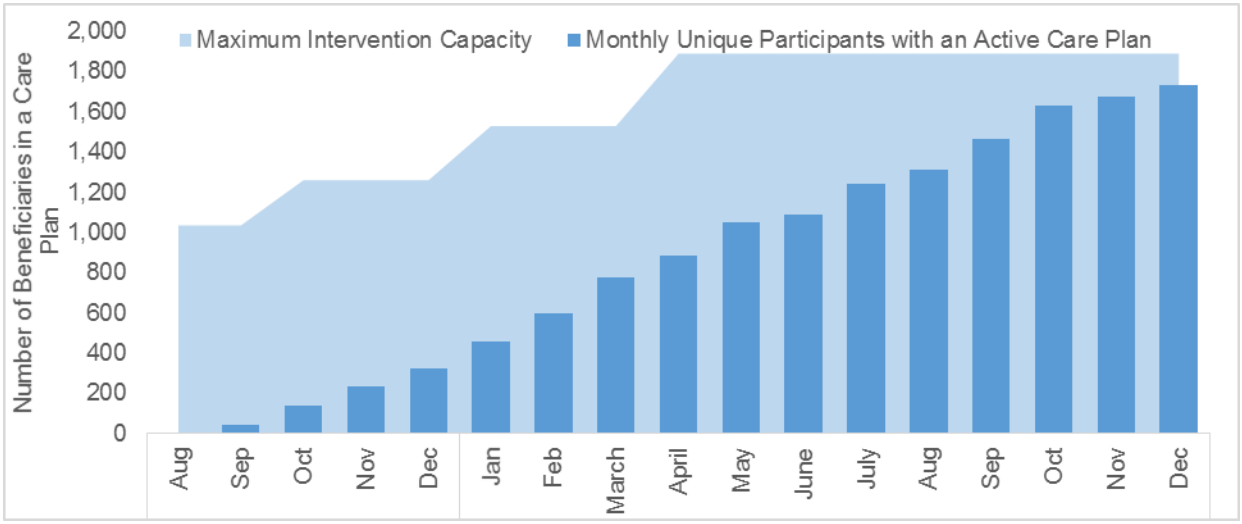
a. Program enrollment

About 38,000 Medicare beneficiaries were attributed to the 14 panels participating in the award. Although all members in this population were eligible to receive services, care plan coordination focused on high-risk participants. The program reported that it had developed care plans for 2,770 Medicare beneficiaries as of March 2015, or roughly 7 percent of all attributed beneficiaries.

b. Service-related measures

Care plans. As of March 2015, 2,770 cumulative participants were or had been enrolled in care plans; the number of participants in active care plans increased throughout the award period (Figure II.A.1). As of April 2015, the maximum capacity target per LCC was reported to be 45 active care plans(although this target has changed over time). Most participants who received a care plan reportedly were in the top band of illness burden scores: 65 percent were in the first band for most advanced illness, whereas 26 percent were in the second band for people with multiple chronic illnesses. The program aims to provide care plans for only a small percentage of high-risk beneficiaries; overall, 15 percent of attributed beneficiaries in the first band received a care plan.

Figure II.A.1. Attributed Medicare beneficiaries in a care plan



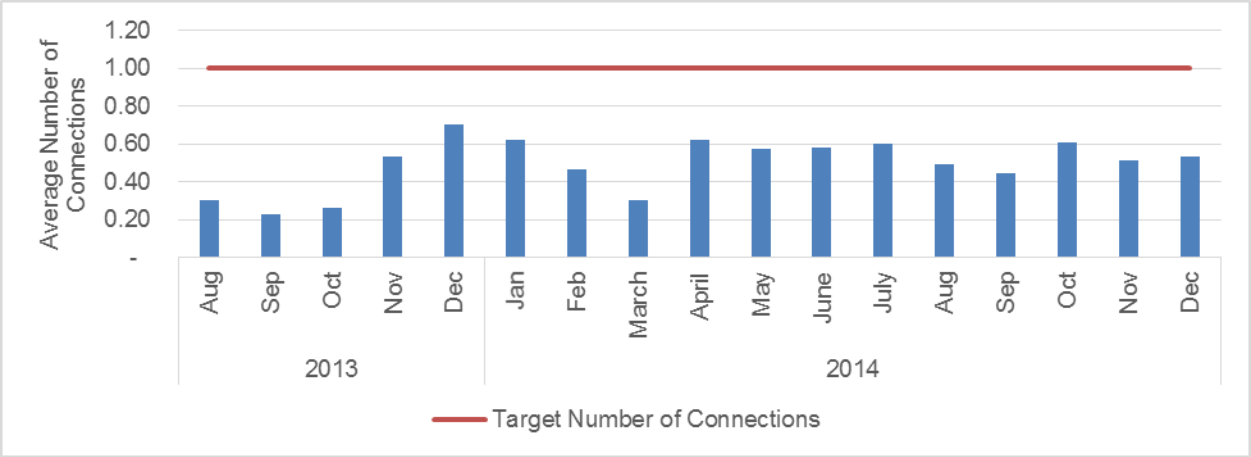
Source: Analysis of CareFirst's HCIA quarterly reports, December 2012 through December 2014, and personal communication with CareFirst, February 2015.

Notes: The number of attributed beneficiaries with care plans is calculated as of December 2014. Maximum intervention capacity is a function of the number of LCCs and the work standard that LCCs are at capacity when they have 45 active care plans.

Care plans can be active for varying lengths of time, depending on whether the participant achieves his or her care plans goals and continues to be engaged in the care planning process. As of December 2014, 54 percent of care plans that had been activated at least 6 months ago were still active (based on a review of care plans that were activated at least six months before the data submission date of December 2014). Among care plans that were still active, the average active time was 10 months (297 days). Care plans that had been closed were active for an average of 6 months (185 days). In April 2015, CareFirst began encouraging providers to close care plans for participants who were either clinically stable or not making progress toward meeting their care plan goals.

LCCs aim to contact participants in an active care plan at least once per week. During the course of the award, LCCs connected with participants on average about once every other week (Figure II.A.2). Most of these connections were via telephone (89 percent), with some in-person visits at the PCPs' offices (8 percent) and a small amount of electronic communication (3 percent). LCCs are not allowed to make home visits to participants.

Figure II.A.2. Average number of LCC contacts per week for participants in an active care plan



Source: Analysis of CareFirst's HCIA quarterly reports, December 2012 through December 2014.

Note: The encounter data reported to Lewin is a count of successful, two-way connections between the LCC and the participant. It does not include administrative calls, messages that the LCC leaves for the participant, conversations between the LCC and specialists, ancillary service providers, one-way emails or text messages to the participants, encounters between providers and participants, and attempted but unsuccessful encounters with the participant.

Participants with an active care plan during 2014 received on average at least one medication reconciliation, though some participants received multiple reviews. About 16 percent of participants in an active care plan received a home-based assessment, 12 percent received enhanced monitoring, and 7 percent used Magellan for behavioral health services.

OIAs. In July 2014, CareFirst paid the first round of OIAs for panels that were able to keep the total cost of care for their attributed Medicare beneficiaries below a specified target. Of the

14 panels, five received an OIA, ranging from $7,843 to $116,045, for a total of payout of $263,003. The award amounts were less than CareFirst initially expected; however, the 2013 performance period was short, reflecting the delayed start of the program.

Panel engagement. CareFirst selected the 14 HCIA-participating panels because they were already highly engaged in its commercial PCMH program. CareFirst believes that individual PCP and overall panel engagement is critical to program success; therefore, the size of OIAs is scaled based on a panel's performance on five quality measures: PCP engagement (35 points), appropriate use of services (20 points), effectiveness of care (20 points), member access (15 points), and structural capabilities (10 points). To measure PCPs' engagement, LCCs submit engagement scores for the participating PCPs, which regional care coordinator supervisors then review and verify. CareFirst assess PCP engagement with the: 1) program; 2) care plan process; 3) beneficiary's satisfaction; 4) analytics capability of the program; and 5) administrative aspects of the program. CareFirst reports that provider engagement with the program increased from 53 percent in the beginning of 2014 to 86 percent at the end of March 2015.

c. Staffing measures

As of December 2014, 158 PCPs had been a part of the HCIA intervention in the 14 participating panels. Through its vendor Healthways, CareFirst hired 44 registered nurse LCCs to help facilitate the care planning process since August 2013. CareFirst also hired five program consultants over the course of the award and recently added four case managers to support the program. CareFirst initially planned to hire 27 LCCs and one program consultant. However, CareFirst hired additional intervention staff, who were approved as part of its first year carry-over funding request, to accelerate implementation after the initial delay and because CareFirst attributed more Medicare beneficiaries to the 14 panels than it had originally projected.

All LCCs completed an initial four-week training class (160 hours), complemented by hands-on experience in the field before beginning their work with panels. LCCs also participate in monthly training forums and weekly update calls, which provide opportunities for continued learning and collaboration. Program consultants received comprehensive training during their first three months. They also had opportunities for continued learning and collaboration.

d. Program time line

CareFirst experienced initial implementation delays due to problems obtaining complete data for participant attribution from CMS. As discussed in the first annual report, CareFirst was initially unable to distinguish primary payer status and beneficiaries were not consistently appearing in the monthly enrollment files (Gilman et al. 2014). CareFirst acquired the necessary data in June 2013 and officially launched the HCIA-funded initiative in August 2013, 13 months later than planned. CareFirst paid out its first round of OIAs in July 2014 for all care delivered in 2013. Though the program was delayed a year, CareFirst reports that it has tried to ramp up the program by hiring more LCCs and intervention staff than originally planned.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research suggests that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.3 introduces each of these facilitators and barriers, which are then described in more detail in the sections that follow.

Table II.A.3. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Adaptability and flexibility of the program • Additional guidance from CareFirst on who could benefit most from a care plan	• Challenges identifying who would benefit the most from a care plan
Implementation process	• PCP engagement • LCCs as a new resource for panels • Using data to identify areas for improvement	• Additional support services not consistently useful to participants • Participant engagement
Internal factors	• Creative methods or incentives to build a positive culture • Prior experience with a similar commercial PCMH program	• No major barriers noted
External factors	• CRISP data system to facilitate identifying eligible participants	• Medical complexity of Medicare patients compared with commercial patients • Technological environment in which some panels operate (for example, EHR functionality and records sharing)

Source: Interviews from second site visit, April 2015; document review, March 2015.

CRISP = Chesapeake Regional Information System for our Patients; EHR = electronic health record.

a. Program characteristics

One continued challenge to program implementation has been consistently identifying which care plan-eligible participants would most benefit from care coordination services. To address this challenge, two characteristics of CareFirst's program facilitated program implementation: (1) PCPs could adopt their own processes for identifying and enrolling participants eligible for care plans, and (2) CareFirst provided additional guidance on who would benefit most from a care plan.

PCPs noted that it is challenging to consistently identify which care plan-eligible participants would benefit most from care coordination services, and CareFirst acknowledges its understanding of how to select the most appropriate beneficiaries for care plans has evolved over time. Although CareFirst sets minimum eligibility guidelines, panels have the freedom to develop their own processes for selecting and enrolling participants into care plans. Many PCPs initially reported using CareFirst's illness burden scores to identify high-risk participants. PCPs and LCCs report that illness burden scores were often too outdated to be useful given delays in

claims data obtained from CMS. In addition, because nearly half of all Medicare beneficiaries fell into the highest-risk tier, illness burden scores were often not considered useful by LCCs and PCPs for narrowing down beneficiaries who would benefit most from a care plan. Over the course of the award, LCCs and PCPs have overcome this challenge by developing their own processes for identifying and enrolling care plan-eligible beneficiaries to supplement the more formal illness burden score criteria initially developed by CareFirst. For example, many PCPs now rely on the Chesapeake Regional Information System (CRISP)—Maryland's statewide health information exchange that provides real-time notifications based on admissions and discharge data to PCPs when their patients were hospitalized—to identify participants who are most appropriate for care plans. The addition of CRISP data allowed LCCs and PCPs to identify beneficiaries with recent hospitalizations, providing additional insight into potential clinical instability and therefore potential appropriateness for care plan enrollment. PCPs also highlighted the importance of using their own clinical judgment when determining who could most benefit from a care plan, because they often know more about a participant's situation or social needs than illness burden scores alone might indicate.

PCPs also had flexibility to determine how to enroll participants into care plans and how to work with LCCs in a way that is most effective and efficient in their existing clinic workflow. For example, some PCPs felt comfortable inviting LCCs into the examination room with a participant to discuss the program, whereas others preferred that LCCs call participants after their visits to invite them to enroll. This flexibility helped PCPs to embrace the program, enabling them to structure it to work best based on their particular practice environment.

Second, although PCPs and LCCs have freedom to select the participants who are most appropriate for care plans, CareFirst has refined the care plan selection process to better target clinically unstable participants, who they consider the major drivers of health care costs. Throughout program implementation, CareFirst learned that some participants can have high illness burden scores, but are not actually clinically unstable. Rather, their high illness burden scores might reflect a recent hospitalization for an acute, non-chronic event. Over time, CareFirst has increasingly emphasized refining its guidance so that LCCs more explicitly focus on clinical instability—which may be evidenced by recent inpatient admissions or ED visits, polypharmacy, or lack of social support—in hopes of preventing costly downstream acute events. CareFirst has also found that social barriers, such as low health literacy and financial concerns, were potential drivers of participants' instability.

b. Implementation process

Three implementation process factors facilitated implementation of CareFirst's program: (1) PCP engagement, (2) the role of the LCCs as an additional care coordination resource for panels, and (3) the use of data to identify areas for improvement. Two additional process factors, (1) additional support services, and (2) participant engagement, demonstrated limited benefit or challenges to program implementation.

Effective implementation of the HCIA-funded initiative depended largely on integrating LCCs into primary care practices, which related directly to having highly engaged PCPs and high quality, engaged LCCs. In addition, using data to identify areas for improvement was a key

implementation facilitator. The availability of additional support services, such as home-based health assessments and in-home monitoring, reportedly enhanced the delivery of the HCIA-funded initiative only minimally.

First, PCP engagement was key to successfully integrating LCCs into their primary care practices and delivering care coordination services to participants. Because CareFirst believes PCP engagement is so important for program success, panels receive engagement scores that affect their annual OIAs. PCPs must be willing to have an LCC based in their practice and engage in the care-planning process. In general, most PCPs in the program were engaged, in part because CareFirst invited only panels that were already highly engaged in the commercial program to participate in the HCIA expansion. CareFirst reported that 95 percent of PCPs activated at least one care plan. The 14 panels participating in the HCIA-funded initiative were among 424 panels enrolled in CareFirst's commercial program. Compared with the 2014 engagement scores of the more than 400 panels not participating in the award, the 14 panels in the HCIA-funded initiative scored 55 percent higher than the average commercial panel.

The second key process factor supporting implementation was the role of the LCCs as an additional care coordination resource for panels. Importantly, the quality of their work was also a consideration for PCPs. PCPs reported that the addition of LCCs was a welcomed resource, as the PCPs would not have had time to focus as intensely on high-risk beneficiaries without program funding to integrate LCCs into the care-planning process. PCPs reported CareFirst provided them with additional LCCs as the number of care plans in their panels increased, so that in general they were not limited by LCC capacity. However, PCPs and regional care coordinators also reported that successful program implementation depended largely on having high quality LCCs. PCPs also reported variation in the quality of LCCs, ranging from those fully integrated into and invested in the care team to others who interacted with PCPs far less frequently and effectively. LCCs are the central players of a beneficiary's care coordination team and receive continuous rigorous training and evaluation. As the program has matured, CareFirst staff report that their knowledge of the necessary skills and characteristics for an effective LCC have increased. As such, they have continued to refine the required skills and abilities for LCCs hired into the program. Site visit respondents reported that in addition to being clinically competent, successful LCCs need to be caring, assertive, self-motivated, communicative, and able to manage their own time. The most successful LCCs had strong personal relationships with their assigned PCPs. Factors that appeared to help build this relationship included LCCs having a presence in the practice(s) as much as possible, their own space to work in the practice, access to the electronic health record (EHR), and a mechanism to educate practice staff that the role of the LCC was to support the PCP and participants with care plans. As the end of the award nears, CareFirst noted that many PCPs will regret losing their LCCs.

Finally, CareFirst used self-monitoring data to inform practice change. Program consultants met quarterly with PCPs to help identify care patterns among their Medicare beneficiaries, aiming to increase quality and cost savings. The consultants reported that these meetings facilitated providers' engagement, with their role shifting over time from illustrating trends to helping improve workflow. For example, in August 2014, program consultants presented to two panels on the importance of glycated hemoglobin, diabetic retinal eye exams, and medical

attention for nephropathy screening rates. The program consultants emphasized that addressing gaps in care was an opportunity to prevent hospitalizations and identify patients who could benefit from care coordination. The PCPs agreed and used CareFirst's participant rosters, in combination with the panels' own EHR data, to identify participants who had not received diabetic screenings. The panel increased its outreach to these participants for diabetic screenings by about 11 percent. However, some PCPs expressed frustration that program consultants seemed to lack the ability to translate data trends into clinically actionable items. For example, even though cancer patients often generate high costs, there are not necessarily relevant action steps that PCPs should be taking as part of the PCMH program to target them directly.

Although CareFirst offered several additional support services for participants with a care plan, PCPs and LCCs did not perceive these services as being consistently beneficial to participants. Panels and PCPs were initially unclear about the availability of these services, which were rolled out at different times and with variable accessibility. Now that they are all available, PCPs reported mixed views on the usefulness of these services. Some PCPs found them moderately helpful, especially the home-based assessments, which enable PCPs to gain a better understanding of what is actually happening in participants' homes. In-home monitoring can also be perceived as helpful to send weight or blood pressure readings daily, but perhaps only for a small proportion of participants and when used appropriately. On the other hand, most PCPs reported that Magellan behavioral health services were difficult to use for locating mental health services for their participants.

Participant engagement is another challenge to delivering care coordination services. The program is delivered almost exclusively through weekly telephone communication between the LCC and participant. A weekly telephone connection can be challenging, particularly at the beginning of the care plan when trust between the participant and the LCC is being built, as well as toward the end of the care plan as the participant becomes more self-reliant. It can sometimes be logistically challenging to connect with participants on the telephone, though LCCs report this is less of an issue than with the younger, commercial population.

c. Internal factors

Two internal factors helped facilitate program implementation: (1) some panels used creative methods or incentives to build a positive culture, and (2) PCPs were able to leverage their knowledge of how the existing commercial PCMH program operated and applied this to the HCIA-funded Medicare expansion. First, some panels used creative methods or incentives to build a positive culture around the program. For example, starting in January 2014, two panels began their own incentive program to reward PCPs $1,000 for every new care plan they activated (funded through their commercial OIA). As a result, these panels had some of the highest engagement scores (both in the HCIA-funded initiative and the commercial program). Other panels enhanced their EHRs so they had additional tools to help them implement the program. Interestingly, CareFirst also anecdotally reported that virtual panels may be more engaged than other panel types, because they have direct control over how they operate their practices and OIAs could have more of a direct impact on their overall practice revenue.

Finally, PCPs reported that it would have been more difficult to implement the HCIA-funded initiative if they had not previously been involved with CareFirst's commercial PCMH program. All panels that participated in the HCIA-funded initiative were already established and functioning, with two to three years of operating experience in CareFirst's commercial PCMH program. CareFirst selected panels that were already engaged with the commercial PCMH program because it believed the prior experience of these panels would allowed PCPs to be better prepared to deal with the more challenging care coordination work required for the Medicare FFS population. CareFirst purposefully minimized the differences between its commercial and Medicare PCMH programs; the commercial program features remained largely intact with minimal modifications to the Medicare program. PCPs reported that it was helpful to build on their knowledge of the commercial program, making the transition rather seamless to extend services to Medicare beneficiaries.

d. External factors

One external factor that facilitated program implementation was the availability of CRISP, the statewide health information exchange that provides real-time notifications to PCPs when their participants are in the hospital. Several panels elected to participate in this initiative, which enabled PCPs to improve transitions of care and identify unstable participants who might benefit from a care plan. Given data lags in CareFirst's illness burden scores for identifying potentially eligible participants for care plans, CRISP has become an increasingly important tool for LCCs and PCPs to identify—in real time—those who could benefit from a care plan. CRISP has been able to provide more actionable insights for both LCCs and case managers participating in the HCIA-funded initiative.

Two external factors presented implementation challenges: (1) the medical complexity of Medicare participants compared with commercial participants and (2) the technological environment in which some panels operate. First, staff reported that it is more difficult and time-consuming to develop care plans for Medicare beneficiaries because they generally have higher rates of chronic disease, are on more medications, and are treated by more specialists. Respondents stated that the relative complexity of Medicare beneficiaries made the development of care plans more difficult. Second, several PCPs (often those participating in virtual panels) reported that they face technological challenges associated with sharing the data necessary for care coordination and management. Although all participating PCPs have access to CareFirst's iCentric platform, they use their own medical records to treat participants because those records contain the most up-to-date and complete information for each participant. Each practice in a panel can have its own electronic or paper-based health record, making information-sharing and collaboration difficult. LCCs report that it is particularly challenging when they do not have access to EHRs, which varied by practice location.

4. Sustainability and scalability

CareFirst believes the PCMH model, and more specifically the HCIA-funded extension to Medicare beneficiaries, is scalable beyond the award period. CareFirst could expand the model for Medicare beneficiaries by inviting more panels to participate in the program and reported that many panels already engaged on the commercial side are eager to have their Medicare

beneficiaries become eligible for the PCMH program. CareFirst also believes the program is financially sustainable by allowing panels, CareFirst, and payers to share in any savings generated by the program. Although the program could theoretically be sustained if several details were developed, such as who would pay for services not currently reimbursed by Medicare, there have been no plans to continue providing services after the award.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

The findings reported above—gathered from interviews with CareFirst program leadership and frontline staff at selected practices during site visit interviews in May 2015—provide important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff were limited to a relatively small number of people involved in implementing CareFirst's HCIA program and could differ from clinicians' views overall. To assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey in fall 2014. Data from the survey provide additional insights into the implementation process and experience, as well as the contextual factors that might affect implementation effectiveness at CareFirst.

In this section, we report the views of clinicians who agreed to participate in CareFirst's HCIA program (which we call the CareFirst sample). First, we focus on the contextual factors that can affect program implementation, including characteristics of the practice locations, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care team functions. We then present data on the alignment of clinicians' views with CareFirst's overall goals for the HCIA program and their views of the barriers to and facilitators of successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

A total of 86 clinicians participating in CareFirst's HCIA program responded to the survey (a response rate of 68 percent). The number of clinicians in each response category does not always sum to 86, here and throughout this section, due to survey item nonresponse, as well as clinicians who reported that a given question did not apply to their practice and thus did not provide a response. In addition, for privacy reasons, data is not included in the tables for survey responses with fewer than 11 respondents. Of these respondents, 76 were physicians and 10 were nurse practitioners. Most of these clinicians practiced at locations with three or more clinicians (67 percent), followed by solo practice locations (24 percent). Most clinicians in the CareFirst sample reported that their primary source of compensation was a salary adjusted for performance (51 percent), followed by fee for service (21 percent), and a fixed salary (18 percent).

The use of health IT is an important pillar of CareFirst's PCMH program. The quality score used to rank panels includes several health IT measures. As shown in Table II.B.1, all clinicians reported using electronic systems for entering clinical notes, and the vast majority reported using it for drug dosing and interaction alerts, prescribing medications, ordering tests and procedures,

and accessing laboratory results. Most clinicians in the CareFirst sample offered patient-facing technologies, including offering their participants the option to do the following online: request a prescription refill, email a clinician about a medical question or concern, and request an appointment.

Table II.B.1. Health IT capacities

Survey item	Number of respondents	Percentage of respondents
The following electronic functionalities are used at the practice location		
Entry of clinical notes	86	100%
Alerts warning of drug dosing or drug interactions	84	98%
Electronic prescribing	82	95%
Access to laboratory test results	80	93%
Ordering of tests and procedures	71	83%
Participant lists or registries	60	70%
Referral tracking	43	50%
Availability of patient-facing technologies		
Participants at this practice location can		
Request refills for prescriptions online	70	81%
Request appointment refills online	61	71%
Email a clinician about a medical question or concern	57	66%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall career, level of burnout, and perceptions of their practice environment can all have an effect on the success of program implementation and organizational change. Overall, clinicians in the CareFirst sample had a high level of satisfaction: for example, 78 percent of clinicians were at least somewhat satisfied with their careers in medicine. About half of clinicians reported only occasionally feeling stressed (48 percent), whereas about one-third experienced some symptoms of burnout when the survey was taken. Most clinicians in the CareFirst sample felt they were supported by their management (76 percent at least somewhat agreed), had adequate opportunities to develop their professional skills (83 percent at least somewhat agreed), and that the amount of work they were expected to finish each week was reasonable (58 percent at least somewhat agreed). Most clinicians in the CareFirst sample believed that most of their work was well-matched to their training (79 percent) and that less than 25 percent of their work could be done by someone with less training (61 percent).

Most clinicians in the CareFirst sample believed that they were able to provide high quality care. As shown in Table II.B.2, 67 percent of clinicians stated that they at least somewhat agreed that they could provide high quality care to all of their participants. The major barriers to providing optimal care reported by clinicians were lack of timely information about care provided to participants by other physicians, time to spend with participants, and sufficient reimbursement.

Table II.B.2. Perceptions of ability to provide high quality care

Survey item	Number of respondents	Percentage of respondents
It is possible to provide high quality care to all of my participants		
Strongly agree	19	22%
Somewhat agree	39	45%
Neither agree nor disagree	12	14%
At least somewhat disagree	15	17%
Percentage reporting each of the following at least somewhat limits their ability to provide optimal, patient-centered care		
I lack timely information about the participants I see who have been cared for by other physicians	73	85%
I do not have enough time to spend with participants during visits	72	84%
The level of reimbursement is not adequate	72	84%
My participants have difficulty paying for needed care	41	48%
I receive too many reminders from my EHR	41	48%
It is difficult for me to obtain specialist referrals for my participants in a timely manner	40	47%
It is difficult for me to obtain specialized diagnostic tests or treatments for my participants in a timely manner	38	44%
I lack adequate information from research evidence to guide my clinical decisions	26	30%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

c. Clinicians' perceptions of care team functioning

It is important to examine clinicians' perceptions of the care team, because a key part of the HCIA program uses LCCs to address participants' needs. LCCs work closely with clinicians to determine which participants could benefit from a care plan, and must communicate closely with clinicians throughout the program. A large majority (80 percent) of clinicians in the CareFirst sample reported working as part of a care team and, overall, their perceptions of how these teams function was positive. Most clinicians surveyed agreed that members of the care team relayed information in a timely manner (88 percent), had sufficient time for participants to ask questions (89 percent), used common terminology when communicating with one another (89 percent), verbally verified information they received from one another (75 percent), and followed a standardized method of sharing information when handing off participants (74 percent).

d. Clinician engagement in other quality improvement activities

Almost two-thirds (64 percent) of clinicians reported participating in quality improvement efforts with other practices, hospitals, government agencies, or professional associations within the past two years. Most clinicians reported received training on quality improvements and tools (64 percent) and conducted at least one clinical audit of care (60 percent).

e. Alignment with goals of HCIA program

Clinicians were asked to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. In Table II.B.3, we present results based on the proportion of clinicians rating each of these goals as extremely important. The

inclusion of the extremely important category helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are simply important. The views of clinicians in the CareFirst sample generally aligned with the goals of CareFirst's PCMH program. Most clinicians in the sample reported that improving care coordination for participants with chronic conditions, reducing hospital readmissions, increasing access to primary care, improving care continuity in primary care, reducing overall health care spending, improving participants' capacity to manage their own care, and reducing emergency department visits were extremely important goals.

Table II.B.3. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Improving care coordination for participants with chronic conditions	55	64%
Reducing hospital readmissions	54	63%
Increasing access to primary care	52	60%
Improving care continuity in primary care	52	60%
Reducing overall health care spending	48	56%
Improving participants' capacity to manage their own care	48	56%
Reducing emergency department visits	46	53%
Improving appropriateness of care	38	44%
Increasing the use of evidence-based practice in clinical care	35	41%
Improving the capability of health care organizations to provide patient-centered care	34	40%
Increasing the number of primary care practices functioning as PCMHs	33	38%
Improving capability of health care organizations to provide team-based care	29	34%
Increasing use of electronic health records and other health IT	26	30%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

3. Awareness of program and perceived effects

Understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on participants' outcomes. For example, if clinicians are aware of the program and believe that it will have a positive effect on the care they provide, they might feel more invested in the program's success. Alternatively, those who feel more negatively about the program could be less likely to enthusiastically implement the HCIA program. In this section, we report on clinicians' experiences with and perceptions of expanding CareFirst's PCMH program to Medicare beneficiaries.

a. Awareness of program and perceived effects

A large majority (88 percent) of clinicians surveyed were at least somewhat familiar with the HCIA program. Among them, overall perceptions of expanding CareFirst's PCMH program to Medicare beneficiaries were positive. Clinicians were familiar with the program were then asked a series of questions about the perceived effect of CareFirst's HCIA program. Most clinicians who were familiar with the program believed it would have a positive effect on the quality of care (68 percent), improve patient-centeredness (75 percent), and have a positive effect on their ability to respond in a timely way to participants' needs (71 percent). Conversely, fewer

than half of clinicians believed it would have a positive effect on safety (47 percent), efficiency (37 percent), and equity (38 percent).Very few clinicians perceived a negative impact of the program; rather, they responded that it was simply too soon to tell the effects of the HCIA program.

b. Barriers to and facilitators of program implementation

The survey also asked clinicians in the CareFirst sample who were at least somewhat familiar with the HCIA program to rate the effect of a series of barriers to and facilitators of program implementation. Of these clinicians, most ranked several factors as having a positive effect on implementation, such as the availability of personnel (68 percent), availability of community resources to care for participants with complex conditions (59 percent), level of program funding (55 percent), availability of relevant participant information at the point of care (55 percent), and the quality of interpersonal communications with other allied health professionals (51 percent). The most often-cited barrier to program implementation was the amount of time required by the program (34 percent) and the amount of required documentation (32 percent). No other barriers were cited by more than 16 percent of responding clinicians.

4. Conclusions about clinicians' attitudes and behavior

Overall we find that several contextual factors helped clinicians implement CareFirst's HCIA program. Clinicians in the CareFirst sample surveyed reported high levels of health IT, low levels of clinician burnout, and generally positive attitudes toward practice management. Most clinicians have a positive perspective towards working with their care teams and believe that they can provide high-quality care to their patients. Most clinicians were aware of CareFirst's HCIA program and believed the program would have a positive effect on quality of care, patient centeredness, and on their ability to respond in a timely way to patient needs. In particular, clinicians believed that the availability of personnel, availability of community resources to care for complex patients, level of program funding, and availability of relevant patient information at the point of care had a positive effect on the program's implementation.

C. Impacts on patient outcomes

1. Introduction

In this section of the report, we draw preliminary conclusions, based on available evidence, about the impacts of CareFirst's HCIA program on patient outcomes in three domains: quality-of-care outcomes, service use, and spending. We first describe the methods for estimating impacts (Section II.C.2) and then the characteristics of the HCIA program panels (also called treatment panels) at the start of the intervention (Section II.C.3). We next demonstrate that the treatment panels were similar at the start of the intervention to the panels we selected as a comparison group, which is essential for limiting potential bias in impact estimates (Section II.C.4). Finally, in Section II.C.5, we describe the quantitative impact estimates, their plausibility given implementation findings, and our conclusions about program impacts in each domain. Our conclusions in this report are preliminary because the analyses do not yet cover the full time over which the intervention is expected to have an effect.

2. Methods

a. Overview

We estimated program impacts as the difference in outcomes between high-risk Medicare beneficiaries assigned to treatment panels and matched comparison panels, adjusting for any pre-intervention differences between the groups. We selected matched comparison panels from the pool of panels participating in CareFirst's commercial PCMH program but not in the HCIA program. This decision reflects our intention to measure the marginal impact of the Medicare HCIA funding, not CareFirst's PCMH program as a whole. It is possible that, before the start of the HCIA program, the commercial program had some positive spillover for Medicare patients. For example, if PCPs developed more cost-effective referral patterns, this might have reduced the total cost of care for all of their patients, not only commercial members. However, any such spillover does not contaminate our impact estimates because we intend to estimate the marginal impact of HCIA funding, separate from any positive spillover effects that might exist without HCIA funding. Further, we anticipate any such spillover to be small, because the primary intervention is individualized care planning for high-risk beneficiaries, and this is likely to have little influence over other patients the panel serves. We limited the sample to high-risk beneficiaries (defined below) because CareFirst's intervention is most likely to affect them and our statistical power to detect true effects is greatest for this population.

In each of the three outcome domains (quality of care, service use, and spending), we specified one or two primary tests before conducting any impact analyses. Each primary test defined an outcome, population, time period, and direction of expected effects for which we hypothesize to see impacts if the program is effective. We drew conclusions about impacts in each domain based on the results of these primary tests and the consistency of the primary test results with the implementation findings and secondary quantitative tests (robustness and model checks).

b. Treatment group definition

The treatment group consists of high-risk Medicare FFS beneficiaries assigned to the 14 treatment panels in four baseline quarters before the intervention began (August 1, 2012, to July 31, 2013) and six intervention quarters (August 1, 2013, to January 31, 2015).

We constructed the treatment group in four steps. First, we used CareFirst's own decision rules to attribute Medicare FFS beneficiaries in each baseline and intervention month to the 14 treatment panels. Specifically, we attributed a beneficiary each month to the PCP (physician or nurse practitioner) who, based on Medicare FFS claims, provided the plurality of primary care services in the past 12 months. If the beneficiary did not have any primary care services in the past 12 months, we attributed him or her to the PCP who provided the plurality of care in the past 24 months. If there was a tie, we attributed to the PCP who provided the most recent service. Then, in each month, we attributed the beneficiary to the treatment panel for which the PCP worked that month. CareFirst provided data on providers who worked in the 14 treatment panels, and when.

Second, in each period (baseline and intervention), we assigned each beneficiary to the first treatment panel he or she was attributed to in the period, and continued to assign him or her to that panel for all quarters in the period. This assignment rule ensures that, during the intervention period, beneficiaries did not exit the treatment group solely because the intervention succeeded in reducing their service use (including visits at treatment panels). The definition for the baseline period corresponds to that of the intervention period so that, across the two periods, interpretation of the population changes over time should be comparable.

Third, we limited the analytic population to those who were at high risk of acute care or other expensive service use since these beneficiaries are the target of CareFirst's care coordination and care transition services. For each baseline quarter, this subgroup consists of the beneficiaries with a Hierarchical Condition Category (HCC) score in the top third among all treatment group members with observable[1] outcomes at the start of the baseline period. The HCC score, developed by the Centers for Medicare & Medicaid Services (CMS), is a continuous variable that predicts a beneficiary's Medicare spending in the following year relative to the national average, with 1.0 indicating that the predicted spending is at the national average and 2.0 indicating that it is twice that average. The HCC score is similar to, but not exactly the same as, the Illness Burden Scores that CareFirst calculates and uses to help identify beneficiaries who would benefit from intensive care coordination services.

In each intervention quarter, the high-risk population consists of beneficiaries whose HCC scores were in the top third among all observable Medicare beneficiaries assigned to the treatment panels at the start of the intervention period.

Fourth, we applied additional restrictions to define the final sample in each quarter. A beneficiary assigned to a treatment panel in a quarter was included in the sample that quarter if he or she (1) had observable outcomes for at least one day in the quarter; (2) lived in Delaware, Maryland, Pennsylvania, Virginia, or Washington, D.C., for at least one day of the quarter; and (3) was not enrolled in Medicaid at any time during the quarter (because CareFirst excludes Medicare–Medicaid dual enrollees from its intervention).

c. Comparison group definition

The comparison group consists of high-risk Medicare FFS beneficiaries assigned to 42 matched comparison panels in each of the baseline and intervention quarters. The comparison panels were similar to the treatment panels during the baseline period on factors that can influence patient outcomes, especially those factors that CareFirst used when deciding which panels to recruit for the intervention. This section describes how we constructed the matched comparison group whereas Section II.C.4 shows the balance we achieved between the two groups on the matching variables.

We identified the 42 comparison panels in four steps. First, at our request, CareFirst provided a list of all 149 panels (of 450) in the commercial program that met the following

[1] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

criteria that all 14 treatment panels also met: (1) located in Maryland, (2) joined the commercial PCMH program when it began in 2011, and (3) served at least 1,000 CareFirst members in 2012. Second, we developed matching variables, defined at the start of the intervention (August 1, 2013), for all treatment and potential comparison panels. These variables include characteristics of the panel overall (for example, the number of PCPs in the panel and the panel's quality and financial performance in the commercial PCMH program); characteristics of all Medicare FFS beneficiaries assigned to the panels (for example, mean HCC score and utilization in the baseline period); and characteristics of high-risk beneficiaries assigned to the panels. When assigning Medicare beneficiaries to the panels, we used the same attribution and panel assignment logic that we used for the treatment panels, as described previously. Section II.C.4 describes the matching variables and their data sources in detail.

Third, we narrowed the pool of 149 to 101 potential comparison panels that, like the treatment panels, (1) had an average of at least 500 assigned Medicare FFS beneficiaries during the four baseline quarters, (2) had at least five PCPs at the start of the intervention, and (3) were located in urban areas.

Finally, we used propensity-score methods to select 42 comparison panels from the pool of 101 that were similar to the 14 treatment panels on the matching variables. The propensity score is the predicted probability, based on all of a panel's matching variables, that a given panel was selected for treatment (Stuart 2010). It collapses all of the matching variables into a single number for each panel that can be used to assess how similar panels are to one another. By matching each treatment panel to one or more comparison panels with similar propensity scores, we generated a comparison group that is similar, on average, to the comparison group on the matching variables. The approach, however, does not ensure that each comparison panel matches exactly to its treatment panel on all matching variables. We prioritized one matching variable—whether a panel is virtual or not—by requiring that a virtual treatment panel could match only to a virtual comparison panel, and a nonvirtual treatment panel could match only to a nonvirtual comparison panel. As noted in Section I, a virtual panel is a group of small, independent practices that agrees to work together to participate in CareFirst's commercial PCMH program. Such panels are likely to have fewer resources, and greater coordination challenges, than the nonvirtual panels, which are part or all of a single, larger practice.

We required each treatment panel to match to at least one, but no more than seven, comparison panels and that the overall ratio of comparison to treatment panels be 3:1. This matching ratio increases the statistical certainty in the impact estimates (relative to a 1:1 overall matching ratio), because it creates a more stable comparison group against which the treatment group's experiences can be compared.

After completing the matching, we assigned Medicare FFS beneficiaries to the comparison practices in each intervention quarter using the same rules we used for the intervention group (see Section II.C.2.b). We also limited the comparison group to high-risk beneficiaries using the same rules as for the treatment group (that is, a beneficiary was in the high-risk group in the intervention quarter if his or her HCC score at the start of the intervention period was in the top

third among all observable Medicare beneficiaries assigned to the treatment panels at the start of the intervention period).d. Construction of outcomes and covariates

We used Medicare claims from August 1, 2009, to January 31, 2015, for beneficiaries assigned to the treatment and comparison panels to develop two types of variables: (1) outcomes, defined for each person in each baseline or intervention quarter; and (2) covariates, which describe a beneficiary's characteristics at the start of the baseline and intervention periods and are used in the regression models for estimating impacts to adjust for beneficiaries' characteristics before the period began. We used covariates defined at the start of each period, without updating them each quarter, to avoid controlling in each intervention quarter for previous quarters' program effects, as this would bias the effect estimates away from detecting true impacts. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. We calculated five quarter-specific outcomes that we grouped into four domains:

1. Domain: Quality-of-care outcomes
    a. Inpatient admissions (number/quarter) for ambulatory care-sensitive conditions (ACSCs);
    b. Number of inpatient admissions followed by an unplanned readmission within 30 days (number/quarter)
2. Domain: Service use
    a. All-cause inpatient admissions (number/quarter)
    b. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
    a. Total Medicare Part A and B spending ($/month)

Four of these outcomes—all but admissions for ACSCs—are outcomes that CMMI has specified as core for the evaluations of all HCIA programs. Our definition of the readmission measure, however, differs from CMMI's standard definition. CMMI typically defines readmissions as the proportion of inpatient admissions that end in an unplanned readmission. Instead, we analyze impacts on the *number* of these unplanned readmissions per thousand beneficiaries per quarter because this enables us to look at the total impact on readmissions across the treatment group, rather than readmissions contingent on an inpatient admission (because the intervention might affect the number and type of admissions as well).

Covariates. The covariates include (1) 18 indicators for whether a beneficiary has each of the following chronic conditions: heart failure, chronic obstructive pulmonary disease, chronic kidney disease, diabetes, Alzheimer's and related dementia, depression, ischemic heart disease, cancer, asthma, hypertension, atrial fibrillation, stroke, hyperlipidemia, hip fracture, osteoporosis, rheumatoid arthritis, bipolar disorder, and schizophrenia); (2) HCC scores; (3)

demographics (age, gender, and race or ethnicity); and (4) original reason for Medicare entitlement (old age, disability, or end-stage renal disease).

e. Regression model

We used a regression model to implement the difference-in-differences design for estimating impacts. For each quarter-specific outcome, the model estimates the relationship between the outcome and a series of predictor variables, assuming that each one of the predictor variables has a linear (additive) relationship with the outcome. The predictor variables include the beneficiary-level covariates (defined in Section II.C.2.d); whether the beneficiary is assigned to a treatment or a comparison panel; an indicator for each practice (which accounts for differences between practices in their patients' outcomes at baseline); indicators for each post-intervention quarter; and an interaction of a beneficiary's treatment status with each post-intervention quarter. The estimated relationship between the interaction term and the outcome in a given quarter is the impact estimate for that quarter. It measures the average difference between outcomes for beneficiaries assigned to the treatment and comparison panels that quarter, subtracting out any differences between these groups during the four baseline quarters. By providing separate impact estimates for each intervention quarter, the model enables the program's impacts to change the longer the panels are enrolled in the program (which is expected to occur). We can also test impacts over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the impact estimates, allowing for statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. The model used robust standard errors to account for clustering of outcomes across quarters for the same beneficiary and a dummy variable for each panel (fixed effects) to implicitly account for clustering of outcomes for beneficiaries assigned to the same panel. Appendix 2 provides details on the regression methods, including descriptions of the weights each beneficiary receives in the model.

f. Primary tests

Table II.C.1 shows the primary tests for CareFirst, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important. The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for detail and a description of how we selected each test). We provided both the awardee and CMMI an opportunity to comment on the primary tests.

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** CareFirst's central goal is to reduce hospitalizations, ED visits, and Medicare Part A and B spending, so our primary tests address these three outcomes. In addition, the primary tests address two quality-of-care outcomes the intervention is expected to affect: hospitalizations for ACSCs and 30-day unplanned hospital readmissions.
- **Time period.** CareFirst expects participating panels to have substantial impacts by their second year of participating in the program, but not in the first. For this reason, our primary tests cover the 5th through 10th quarters (I5 through I10) after the intervention began, which

Table II.C.1. Specification of the primary tests for CareFirst Blue Cross Blue Shield

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population	Substantive threshold (impact as percentage of the counterfactual)[c,d]
Quality-of-care outcomes (2)	Inpatient admissions for ambulatory care-sensitive conditions (#/person/quarter)	Average over intervention quarters 5 through 10	High-risk Medicare FFS beneficiaries assigned to treatment panels	-15.0
	30-day unplanned hospital readmissions (#/person/quarter)			-7.7
Service use (2)	All-cause inpatient admissions (#/person/quarter)			-7.7
	Outpatient ED visit rate (#/person/quarter)			-7.7
Spending (1)	Medicare Part A and B spending ($/person/month)			-6.2

Notes: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

High-risk beneficiaries are defined as those with an HCC score in the top third at the start of the baseline or intervention period, as described in the text.

[a] We adjusted the *p*-values from the primary test results for the multiple comparisons made within each domain, but not across domains.

[b] The regression models for estimating program impacts controlled for differences in outcomes between the pre-intervention treatment and comparison groups.

[c] For all but one outcome, we set the substantive threshold to 75 percent of our calculation of CareFirst's expected effects for high-risk beneficiaries during the primary test period. For hospitalizations for ambulatory-care sensitive conditions, for which CareFirst did not set an explicit target, we used the reductions in acute care that Peikes et al. (2011) indicated could be feasible among high-risk beneficiaries in a patient-centered medical home program.

[d] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention.

ED = emergency department; FFS = fee-for-service; HCC = Hierarchical Condition Category.

corresponds to the period from August 2014 through January 2016. We include one month beyond when CareFirst is currently set to end (December 2015, given its recent no-cost extension) so that we can include outcomes for the quarter that runs from November 2015 to January 2016, most of which falls when the program will still be operating.

- **Population.** We chose high-risk beneficiaries (as defined in Section II.C.2.b) for the primary tests because (1) CareFirst expects its overall impacts to be driven by impacts on high-risk beneficiaries, whom CareFirst targets for intensive care management services; and (2) the statistical power to detect effects is greatest for this group.
- **Direction (sign) of the impact estimate.** The primary tests are testing for a reduction relative to the counterfactual, for each of the outcome measures.

  Substantive thresholds. Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant; for this reason, we have pre-specified thresholds for what we call substantive importance. We express the threshold as a percentage change from the counterfactual—that is, the outcomes that beneficiaries in the treatment group would have had if they had not received the treatment. For the high-risk subgroup, the 6.2 to 7.7 percent thresholds we chose (depending on the outcome) are 75 percent of our calculation of CareFirst's expected effects during the primary test period (intervention quarters 5 through 10). (We use 75 percent recognizing that CareFirst could still be considered successful if it approached, but did not achieve, its fully anticipated effects.) The 15 percent threshold for ACSC hospitalizations is extrapolated from the literature (Peikes et al. 2011) because CareFirst did not specify by how much it expected to reduce these hospitalizations.

Due to limitations in data availability, we were able to conduct the primary tests in this report only partially. Specifically, we estimated impacts only through the 5th and 6th intervention quarters (August 2014 through January 2015), and did not include quarters 7 through 10. Future reports will cover the full 18 months from quarters 5 through 10.

g. Secondary tests

We also conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the non-experimental design or random fluctuations in the data. We will have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. Specifically, we estimated the program's impacts on hospitalizations and total Medicare spending for the high-risk beneficiaries during two additional intervention periods: (1) the first 6 months after the panels joined the intervention (intervention quarters 1 and 2), and (2) months 7 to 12 (quarters 3 and 4). Because we and CareFirst expect program impacts to increase over time, with little or no impacts in the first few months of the program, the following pattern would be highly consistent with an effective program—little to no measured effects in the first two quarters, growing effects in quarters 3 and 4, and the largest impacts in quarters 5 and 6 (which is the period for the primary tests covered in this report). In contrast, if we found very large differences in outcomes

(favorable or unfavorable) in the first 6 intervention months, this could suggest a limitation in the comparison group, not true program impacts.

h. Synthesizing evidence to draw conclusions

Within each domain, we drew one of four conclusions about program effectiveness, based on the primary test results, the results of secondary tests, and the plausibility of those findings given the implementation evidence. These four possible conclusions are as follows: (1) statistically significant favorable effect (the highest level of evidence), (2) substantively important favorable effect, (3) indeterminate effect, and (4) substantively important unfavorable effect. (We cannot conclude that a program has a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program harms.)

Our decision rules for each of the four possible conclusions are described in Appendix 3. In short, we concluded that a program had a statistically significant favorable effect in a domain if (1) at least one primary test result in the domain was favorable and statistically significant, after adjusting the statistical tests to account for multiple tests (if applicable) within a domain; or (2) the average impact estimate across all primary tests in the domain was favorable and statistically significant. In both cases, we also need to determine that the primary test results were plausible given the secondary tests and implementation evidence. We concluded that a program had a substantively important favorable effect if the average impact estimate in the domain was substantively important but not statistically significant, and if the result was plausible given the secondary tests and implementation evidence. In contrast, if the average impact estimate was unfavorable (opposite the hypothesized direction), larger than the substantive threshold, and unfavorable effects are plausible given the other evidence, we concluded the program had a substantively important unfavorable effect. Finally, if the tests in a domain did not meet any of these criteria, we concluded that the impact in that domain is indeterminate.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the treatment group at the start of the intervention (August 1, 2013), which can be seen in the second column of Table II.C.2. (Table II.C.2 also serves a second purpose—to show the equivalence of the treatment and comparison panels at the start of the intervention—which we describe in Section II.C.4.)

Characteristics of the panels overall. At the start of the intervention, the 14 treatment panels, on average, consisted of nine PCPs. Half of the panels were virtual, meaning they consisted of several small practices that joined together contractually to participate in CareFirst's commercial PCMH program. This proportion is consistent with CareFirst's overall commercial program, in which about half of the 450 panels are virtual. Two of the 14 treatment panels were owned by health systems, again consistent with the proportion (15 percent) of panels that are of this type in the commercial program. The treatment panels performed well in the commercial program in 2011 and 2012, achieving an average 4 percent savings against expected 2011–2012 care costs and an average quality score over those two years of 68 out of 100. In contrast, the average savings across the 101 panels in the potential comparison pool was 2 percent, and the

mean quality score was 64. (Table II.C.2 shows characteristics of the 101 panels in our refined comparison pool because comparable data are not available for all 450 panels in the commercial program). The treatment panels practiced in relatively affluent zip codes, where the mean household income was almost $78,000 from 2008 to 2012 (compared with a national average of $53,046).

Characteristics of the panels' Medicare FFS beneficiaries. The characteristics of all Medicare nondual FFS beneficiaries assigned to the treatment panels during the baseline period (August 1, 2012, through July 31, 2013) were, overall, similar to the nationwide FFS averages. The HCC risk score for the treatment group (1.1) was just about the national average (1.0). Patients in the treatment panels also had hospital admission rates, total Medicare spending, and 30-day readmission rates that were close to the national averages. The mean outpatient ED visit rate (81/1,000 people/quarter) was lower than the national average of 105, which could in part be due to the fact that the treatment group excludes those dually enroll in Medicare and Medicaid, who often have high outpatient ED visit rates (Congressional Budget Office 2013).

The high-risk Medicare FFS beneficiaries assigned to the panels had substantially greater health care needs during the baseline period than the full treatment group. Their mean HCC risk score was about twice the mean for all treatment group members (2.0 versus 1.1). Further, they had approximately twice the number of all-cause inpatient admissions and Medicare spending, 70 percent more outpatient ED visits, and 20 percent higher 30-day unplanned hospital readmission rates. These comparisons are between the high-risk patients and all patients (which includes high-risk patients); differences would be even larger if we compared the high-risk patients to those who are not at high risk.

Consistent with CareFirst's plan to target high-risk beneficiaries for care plans and the intensive care management services that follow, the beneficiaries we identified as high risk were almost five times more likely to have received a care plan as of December 31, 2014, than those not in the high-risk group (results not shown). Specifically, based on data from CareFirst, which we merged with the treatment group beneficiaries, 13 percent of Medicare FFS beneficiaries in the high-risk group during any of the intervention quarters had received a care plan as of December 31, 2014, whereas only 3 percent of those not in the high-risk group did.

Table II.C.2. Characteristics of treatment and comparison panels before the intervention start date (August 1, 2013)

Characteristic of panel	Treatment panels (N = 14)	Unmatched comparison pool (N = 101)	Matched comparison group (N = 42)	Absolute difference[a]	Standardized difference[b]	Medicare FFS national average
Exact match variable[c]						
Characteristics of the panel overall						
Panel type: Virtual (%)	50.0	56.4	50.0	0	0	n.a.
Propensity-matched variables[d]						
Characteristics of the panel overall						
Average quality score for the commercial program in 2011 and 2012[e]	68.1	64.3	66.4	1.64	0.239	n.a.
Average cost savings in the commercial program in 2011 and 2012 (%)[f]	3.9	2.1	3.2	0.7	0.190	n.a.
PCPs in panel who work in practices that are medical homes (%)	34.7	12.2	29.6	5.1	0.156	n.a.
Panel type: Health system (%)	14.3	10.9	8.2	6.1	0.237	n.a.
Number of PCPs	9.29	8.73	8.53	0.76	0.263	n.a.
Characteristics of a panel's practice(s) location(s)						
Median household income in zip code(s) where panel's practice(s) are located ($)	77,982	77,203	78,406	-424	-0.020	53,046[g]
Characteristics of all Medicare FFS, nondual patients assigned to panels during the baseline year (August 1, 2012 – July 31, 2013)						
Number of beneficiaries	2,202	1,538	1352	850**	1.208	n.a.
HCC risk score	1.08	1.09	1.07	0.01	0.082	1.0
All-cause inpatient admissions (#/1,000 patients/quarter)	79.87	78.56	79.22	0.65	0.044	74[h]
Outpatient ED visit rate (#/1,000 patients/quarter)	81.33	94.53	82.66	-1.33	-0.082	105[i]
Medicare Part A and B spending ($/patient/month)	998	995	988	10	0.073	860[j]
30-day unplanned hospital readmission rate (%)	15.4	16.2	15.7	-0.3	-0.108	16.0[k]
30-day unplanned hospital readmissions (#/person/quarter)[l]	10.96	11.24	10.81	0.16	0.047	n.a.
Inpatient admissions for ambulatory care-sensitive conditions (#/1,000 person/quarter)[l]	13.28	13.37	12.92	0.36	0.094	11.8[m]
Disability as original reason for Medicare entitlement (%)	11.2	12.4	10.8	0.4	0.111	16.7[n]
Age (years)	73.84	73.53	73.87	-0.03	-0.022	71[o]
Female (%)	59.2	58.9	58.7	0.5	0.137	55.3[n]
Race: White (%)	85.1	77.1	82.0	3.2	0.207	81.8[n]

Table II.C.2 *(continued)*

Characteristic of panel	Treatment panels (N = 14)	Unmatched comparison pool (N = 101)	Matched compar-ison group (N = 42)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Characteristics of high-risk Medicare FFS, nondual patients assigned to panels during the baseline year (August 1, 2012 – July 31, 2013)						
Number of high-risk beneficiaries	693	498	427	266**	1.043	n.a.
HCC risk score	2.00	2.01	2.00	0.01	0.084	1.0
All-cause inpatient admissions (#/1,000 patients/quarter)	160.58	155.39	157.88	2.70	0.127	74
Outpatient ED visit rate (#/1,000 patients/quarter)	136.44	153.47	139.26	-2.82	-0.103	105
Medicare Part A and B spending ($/patient/month)	1,843	1,843	1,832	11	0.050	860
30-day unplanned hospital readmission rate (%)	18.3	19.4	18.2	0.1	0.031	16.0
30-day unplanned hospital readmissions (#/person/quarter)[l]	25.96	26.50	25.16	0.80	0.110	n.a.
Inpatient admissions for ambulatory care-sensitive conditions (#/person/quarter)[l]	32.29	30.92	30.57	1.73	0.254	11.8
Omnibus test for balance on matching variables[p]						
p-value			0.39			n.a.

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code household income data merged from the American Community Survey ZIP Code Characteristics. CareFirst provided data on characteristics of the panels, including quality scores and financial performance in the commercial program.

Notes: The comparison group means are weighted based on the number of matched comparison panels per treatment panel. For example, if four comparison panels are matched to one treatment panel, each of the four comparison panels has a matching weight of 0.25.

Absolute differences might not be exact due to rounding.

We did not audit or independently confirm the quality or financial performance scores that CareFirst reported for the panels in the commercial medical home program.

[a] The absolute difference is the difference in means between the treatment and matched comparison groups.

[b] The standardized difference is the difference in means between the treatment and matched comparison groups divided by the standard deviation of the variable. The standard deviation is calculated among the pooled treatment and matched comparison groups.

[c] Exact match means that a virtual treatment panel could be matched only to a virtual comparison panel, and a nonvirtual treatment panel could be matched only to a nonvirtual comparison panel.

[d] Variables that we matched on through a propensity score, which capture the relationship between a panel's characteristics and its likelihood of being in the treatment group.

[e] Average quality score for CareFirst's commercial program for 2011 and 2012. The quality score is out of 100 points.

[f] Average financial performance in the commercial program is a function of credits (global projected care costs) minus debits (all services paid) for 2011 and 2012.

[g] U.S. Census Bureau, 2008–2012 American Community Survey, Median household income.

[h] Chronic Conditions Data Warehouse (2014b).

[i] Gerhardt et al. (2014).

[j] Boards of Trustees (2013).

[k] Centers for Medicare & Medicaid Services (2014).

[l] These measure are included on the table for descriptive purposes but were not included in the matching model.

[m] This rate is for individuals ages 65 and above (Truven Health Analytics 2015).

Table II.C.2 *(continued)*

[n] Chronic Conditions Data Warehouse (2014a, Table A.1).

[o] Health Indicators Warehouse (2014a).

[p] Results from an overall chi-square test indicate the likelihood of observing differences in the matching variables as large as the differences we observed if, in fact, the treatment and comparison populations (from which we drew the samples) were perfectly balanced. The value of $p = 0.39$ for the chi-square test suggests that the two groups are well balanced, because we cannot reject the null hypothesis that their characteristics are identical.

*/**/*** Significantly different from zero at the .10/.05/.01 level, two-tailed test, respectively. No differences were significantly different from zero at the 0.01 level.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; HCC = Hierarchical Condition Category; SD = standard deviation; PCP = primary care provider.

n.a. = not applicable.

4. Equivalence of the treatment and comparison groups at the start of the intervention

Demonstrating that the treatment and comparison groups are similar at the start of the intervention is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying difference-in-differences models—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment group not received the intervention.

Table II.C.2 shows that the 14 treatment panels and the 42 selected comparison panels were similar at the start of the intervention on most matching variables. By construction, there were no differences between the two groups on the exact matching variable—whether the panel was virtual. There were some differences between treatment group beneficiaries and matched comparison group beneficiaries on the variables we matched through propensity scores, but the standardized differences across the propensity-score matching variables are almost all within our target of 0.25 standardized differences, and most were within 0.15 standardized differences (the 0.25 target is an industry standard; for example, see Institute of Education Sciences 2014). The omnibus test that the treatment and comparison panels are perfectly matched on all variables cannot be rejected ($p = 0.39$), further supporting the premise that the treatment and comparison groups were similar at the start of the intervention.

The propensity matching technique improved or did not affect the balance for most variables relative to the unmatched comparison pool, but worsened the balance for a few. This can be seen in Table II.C.2, which shows the means for the full comparison pool and for the selected comparison group. Specifically, propensity matching improved balance on a panel's quality and financial performance in the commercial program, which makes sense because CareFirst recruited top performers for the HCIA program but the potential comparison pool was not restricted to top performers. Propensity matching also improved balance on whether the panel's practices were certified as PCMHs at baseline and the percentage of beneficiaries in panel's practices who were non-Hispanic whites. The improvements in balance on some variables came at the expense of increasing the differences between the treatment and comparison panels in the mean number of (1) primary care providers and (2) assigned beneficiaries in the baseline period.

On average, the treatment panels had slightly more PCPs (by 0.76 providers) and considerably more attributed Medicare FFS beneficiaries, overall (by 850 beneficiaries) and for the high-risk participants (by 266). However, in discussion with CMMI, we determined that—

although these two variables fell outside our preferred standard—it is reasonable to accept the selected comparison group for three reasons. First, we can account for differences in panel size through regression weights in our impact analyses. Second, there is no correlation between the number of attributed beneficiaries and the outcomes during the baseline period (results not shown), so differences in size within the observed range are unlikely to bias the impact results. Third, if there were any systematic differences in outcomes (that do not vary over time) that result from a different number of primary care providers or beneficiaries, the difference-in-differences model would account for them.

Overall, the propensity matching had little impact on the balance for most characteristics of a panels' assigned beneficiaries, reflecting the fact that the comparison pool was already narrowed to a group that looked similar to the treatment panels on these dimensions

5. Intervention impacts

In this section, we first present sample sizes and mean outcomes, by quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates that follow; however, the differences in mean outcomes are not regression-adjusted and not impact estimates by themselves. Next, we present the results of the primary tests, by domain. Then, we present the secondary tests results and assess whether the primary test results are plausible given the secondary tests and whether primary test results are plausible given the implementation evidence. We end with preliminary conclusions about program impacts in each domain.

a. Sample sizes

In the first baseline quarter (B1), the treatment group included 9,665 beneficiaries assigned to 14 panels and the comparison group included 20,172 beneficiaries assigned to 42 panels (see Table II.C.3). By construction, these groups—which are limited to high-risk beneficiaries—were one-third of all Medicare FFS beneficiaries assigned to the panels. The sample sizes stayed relatively steady across the baseline and intervention quarters, reflecting near balance of two opposing forces—beneficiaries being added to the sample because they are newly assigned to the panels and beneficiaries dropping out of the sample because they die, move from the region, switch from FFS to managed care, or enroll in Medicaid. As expected, the sum of the comparison group members' weights was roughly equal to the size of the treatment group in each baseline quarter.

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

Quality-of-care outcomes. For both the treatment and comparison groups, the number of hospitalizations for ACSCs declined moderately in the intervention period. The differences between the groups were small (less than 5 percent) in all quarters, and not consistently positive or negative. The 30-day unplanned readmission rates (number per quarter) were also very similar for the treatment and comparison groups in all quarters.

Table II.C.3. Sample sizes and unadjusted mean outcomes for high-risk Medicare FFS beneficiaries in the treatment and comparison groups for CareFirst, by quarter

	Number of Medicare FFS beneficiaries (panels)			Inpatient admissions for ambulatory care-sensitive conditions (#/1,000/quarter)			30-day unplanned hospital readmissions (#/1,000/quarter)			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/month)		
Q	T	C (no wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (August 1, 2012 – July 31, 2013)																		
B1	9,665 (14)	20,172 (42)	9,680	30.2	34.5	-4.3 (-12.4%)	26.1	27.3	-1.2 (-4.6%)	161.1	159.7	1.4 (0.9%)	136.9	146.3	-9.4 (-6.4%)	$1,870	$1,850	$20 (1.1%)
B2	9,686 (14)	20,383 (42)	9,770	33.2	29.9	3.3 (11.1%)	25.0	25.2	-0.2 (-0.7%)	163.8	158.7	5.2 (3.3%)	128.6	132.9	-4.3 (-3.2%)	$1,750	$1,816	-$66 (-3.6%)
B3	9,793 (14)	20,146 (42)	9,669	34.7	33.5	1.2 (3.5%)	26.0	23.5	2.5 (10.7%)	164.0	160.1	3.9 (2.4%)	125.0	122.7	2.3 (1.9%)	$1,858	$1,824	-$34 (1.9%)
B4	9,673 (14)	20,017 (42)	9,698	28.1	25.6	2.5 (9.8%)	25.5	22.2	3.4 (15.2%)	156.2	148.3	7.9 (5.3%)	140.4	145.7	-5.3 (-3.7%)	$1,877	$1,766	111 (6.3%)
Intervention period (August 1, 2013 – January 31, 2015)																		
I1	10,550 (14)	21,011 (42)	10,488	25.8	26.9	-1.2 (-4.3%)	27.1	22.2	5.0 (22.4%)	153.7	145.0	8.8 (6.1%)	132.1	128.9	3.2 (2.5%)	$1,904	$1,836	68 (3.7%)
I2	10,539 (14)	21,253 (42)	10,532	29.0	28.4	0.6 (2.2%)	25.7	23.9	1.9 (7.8%)	148.0	143.2	4.8 (3.3%)	125.0	119.7	5.2 (4.4%)	$1,701	$1,694	7 (0.4%)
I3	10,337 (14)	21,232 (42)	10,470	27.5	26.4	1.0 (4.0%)	24.6	18.3	6.3 (34.5%)	147.2	139.3	7.9 (5.7%)	132.4	126.1	6.3 (5.0%)	$1,792	$1,701	92 (5.3%)
I4	10,289 (14)	21,255 (42)	10,461	29.2	25.3	3.8 (15.1%)	22.6	16.0	6.6 (41.1%)	147.7	136.5	11.2 (8.2%)	146.6	144.7	1.8 (1.3%)	$1,756	$1,702	53 (3.2%)
I5	10,202 (14)	21,528 (42)	10,471	28.2	26.7	1.5 (5.7%)	25.2	20.5	4.7 (22.7%)	143.2	140.4	2.8 (2.0%)	134.0	144.2	-10.3 (-7.1%)	$1,786	$1,799	-12 (-0.7%)
I6	10,125 (14)	21,373 (42)	10,400	28.9	31.8	-2.8 (-8.9%)	25.1	25.0	0.1 (0.5%)	158.1	153.0	5.2 (3.4%)	140.2	137.8	2.3 (1.7%)	$1,764	$1,708	56 (3.3%)

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The baseline quarters are measured relative to the start of the baseline period on August 1, 2012. For example, the first baseline quarter (B1) runs from August 1, 2012, to October 31, 2012. The intervention quarters are measured relative to the start of the intervention period on August 1, 2013. For example, the first intervention quarter (I1) runs from August 1, 2013 to October 31, 2013. In each period (baseline or intervention), the treatment group each quarter includes all high-risk beneficiaries who were assigned to a treatment panel by the start of the quarter and who met other sample criteria—that is, they were enrolled in FFS Medicare, were living in Maryland or surrounding areas, and were not enrolled in Medicaid. In each period, the comparison group includes all high-risk beneficiaries who were assigned to a comparison panel by the start of the quarter and who met the other sample criteria. See text for details.

Table II.C.3 *(continued)*

The outcome means were weighted such that (1) each treatment beneficiary gets a weight of 1; and (2) each comparison beneficiary gets a weight that is the product of two weights: (1) a matching weight, equal to the reciprocal of the total number of comparison panels matched to the same treatment panel as the beneficiary's assigned panel, and (2) a practice size weight, which equals the average number of high-risk beneficiaries assigned to the matched treatment panel during the four baseline quarters divided by the average number of high-risk beneficiaries assigned to the beneficiary's comparison panel over those quarters. The difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; Diff = difference; ED = emergency department; FFS = fee-for-service; I = intervention; Q = quarter; T = treatment; no wgt = unweighted; wgt = weighted.

NA = not available.

n.a. = not applicable.

Service use. The hospitalization rates for both the treatment and comparison groups declined steadily from B3 to I5 (by 10 to 15 percent), before increasing in the last intervention quarter. The hospitalization rates were modestly higher (0.9 to 6.1 percent higher) for the treatment group than the comparison groups in all quarters, without any consistent trend of increasing or decreasing differences.

Spending. The mean Medicare Part A and B spending for the comparison group was similar to the treatment group (within 6.3 percent) for all baseline and intervention quarters.

c. Results for primary tests, by domain

Overview. The primary tests reflect the average impact of the intervention in the second and third year of the intervention. For this report, we had data available only for the first six months of this period. Thus, the primary tests in this report reflect impacts over only two intervention quarters (I5 and I6). For each of the five outcomes in the three domains, the regression-adjusted differences between the treatment and comparison groups during the two quarters of the primary test period were small (see Table II.C.4). None of these differences were statistically significant or larger than the substantive thresholds in either a favorable or unfavorable direction.

Quality-of-care outcomes. The rate of ACSC hospitalizations for the treatment group during the primary test period was 3.8 percent lower than our estimate of the counterfactual, and the rate of unplanned readmissions was 4.8 percent higher. (Our estimate of the counterfactual is the treatment group mean minus the difference-in-differences estimate.) Neither difference was statistically significant or substantively large. After combining results across the two outcomes in this domain, the outcomes for the treatment group were almost identical (0.3 percent higher) to the outcomes for the estimated counterfactual.

The statistical power to detect effects was marginal for ACSC hospitalizations, but poor for 30-day unplanned readmissions. For example, Table II.C.4 indicates that the tests had a 62.8 percent likelihood of detecting an effect on ACSC hospitalizations that was, in truth, the size of the substantive threshold. Power was worse (24.6 percent) for readmissions because of the smaller substantive threshold and greater variation in the outcome.

Service use. The treatment group's average hospitalization rate was 0.9 percent lower, and the outpatient ED visit rate was 1.3 percent higher, than the estimated counterfactual. Neither of these differences was statistically significant or substantively large. After combining results across the two outcomes in this domain, the outcomes for the treatment group were almost identical (0.2 percent higher) to the outcomes for the counterfactual. Power to detect effects that were the size of the substantive thresholds was marginal (63.3 and 56.8 percent, respectively) for the individual outcomes but good (75.2 percent) for the two outcomes combined.

Spending. The treatment group averaged $1,775 in Part A and B spending ($/person/month) during the primary test period, which was 1.5 percent (or $27) lower than the estimated counterfactual. However, this difference was neither statistically significant ($p = 0.36$) nor close to the substantive threshold of 6.2 percent. Statistical power to detect an effect the size of the substantive threshold was, again, marginal (58.2 percent).

Table II.C.4. Results of primary tests for CareFirst

Primary test definition					Statistical power to detect an effect that is[b]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual[a])	Size of the substantive threshold	Twice the substantive threshold[c]	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual [a] (standard error)	Percentage difference[d]	*p*-value[e]
Quality-of-care outcomes (2)	Inpatient admissions for ambulatory care-sensitive conditions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 5–6 (August 1, 2014 to January 31, 2015)	All observable[f] high-risk Medicare FFS beneficiaries assigned to treatment panels	-15.0%	62.8%	97.3%	28.6	-1.1 (2.8)	-3.8%	0.445[g]
	30-day unplanned readmissions (#/1,000/quarter)			-7.7%	24.6%	46.4%	25.1	1.1 (3.1)	4.4%	0.539[g]
	Combined (%)			-11.4%	46.9%	87.0%	n.a.	n.a.	0.3%	0.512[h]
Service use (2)	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)			-7.7%	63.3%	97.5%	150.7	-1.4 (7.2)	-0.9%	0.488[g]
	Outpatient ED visits (#/1,000/quarter)			-7.7%	56.8%	94.8%	137.1	1.7 (7.2)	1.3%	0.518[g]
	Combined (%)			-7.7%	75.2%	99.6%	n.a.	n.a.	0.2%	0 .517[h]
Spending (1)	Medicare Part A and B spending ($/beneficiary/month)			-6.2%	58.2%	95.5%	$1,775	-$26.7 (75.1)	-1.5%	0.361

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[b] The power calculation is based on actual standard errors from the analysis. For example, in the last row, a 6.2 percent effect on Medicare Part A and B spending (from the counterfactual of $1,775 +$26.70 = $1,801.70) would be a change of $112. Given the standard error of $75.10 from the regression model, we would be able to detect a statistically significant result 58.2 percent of the time if the impact was truly $112, assuming a one-sided statistical test at the $p = 0.10$ significance level.

Table II.C.4 *(continued)*

[c] We show statistical power to detect a very large effect (twice the size of the substantive threshold) because this provides additional information about the likelihood that we will find effects if the program is indeed effective. If power to detect effects is less than 75 percent even for a very large effect, then the evaluation is extremely poorly powered for that outcome.

[d] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison group, divided by the adjusted comparison group mean.

[e] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test.

[f] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

[g] We adjusted the *p*-values from the primary test results for the multiple (two) comparisons made within the service use domain, and (separately) for the two comparisons made within the quality-of-care outcomes domain.

[h] This *p*-value tests the null hypothesis that the difference-in-differences estimates across the two outcomes in the domain, each expressed as percentage change from the estimated counterfactual, is greater than or equal to zero (a one-sided test).

ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

d. Results for secondary tests

As shown in Table II.C.5, the differences in hospitalizations and spending for the treatment group and its estimated counterfactual were small (less than 2.5 percent) and not statistically significant during the two secondary test periods: the first six months of the intervention (I1 and I2) and the next six months (I3 and I4). These results help support the credibility of the comparison group because we do not see large differences (favorable or unfavorable) during the first year of panel participation, a period during which we and the awardee did not expect to see large program effects. This increased confidence in the comparison group, in turn, gives us greater confidence in the primary test results.

e. Consistency of quantitative estimates with implementation findings

The impact estimates in the primary tests are plausible given the implementation findings. The primary tests did not find any effects (favorable or unfavorable) during the first six months of the primary tests period that were statistically significant or substantively important. The implementation evidence shows the program was active during these six months. For example, as described in Section II.A.2.b, care managers provided intensive care management services to 1,300 to 1,800 high-risk Medicare beneficiaries during this period. Therefore, the lack of measured effects is not simply due to the program failing to deliver a meaningful intervention. However, even with a well-implemented intervention, it is possible that the program was not able to change participants' or providers' behaviors in ways that would affect study outcomes during the relatively short part of the primary test period covered in this report (6 of a planned 18 months).

f. Conclusions about program impacts, by domain

Based on all evidence currently available, we have drawn the preliminary conclusion that the program impact *is indeterminate in each of the three domains:* quality-of-care outcomes, service use, and spending. These conclusions are summarized in Table II.C.6. We reached these conclusions because (1) in each domain, the primary test results were neither statistically significant nor substantively large; (2) the secondary tests helped to confirm the credibility of the comparison group used in the primary tests, by showing that there were no estimated effects in the first year of program operations—a period when we and the awardee expected little or no effects; and (3) the results are plausible given the implementation evidence.

These conclusions have different implications depending on the outcome domain. For the service use domain, the statistical power to detect effects at least as large as the substantive threshold was good (for the test combining hospital admissions and outpatient ED visits). Therefore, although the program might have had a small effect, it likely did not have a substantively large effect for the study population over the period examined. In contrast, for the other two domains (quality of care and spending), the power to detect effects was marginal or poor. Therefore, the lack of measured effects could mean the program (1) did not have substantively large effects in these domains; or (2) it did, but our statistical tests failed to detect them.

Table II.C.5. Results of secondary tests for CareFirst

Secondary test definition				Results			
Domain	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between treatment and the estimated counterfactual (standard error)	Percentage difference[a]	*p*-value[b]
Service use	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)	Intervention quarters 1, 2	All observable high-risk Medicare FFS beneficiaries attributed to treatment panels	150.9	1.0 (6.9)	0.7%	0.557
		Intervention quarters 3, 4		147.5	3.2 (7.1)	2.2%	0.676
Spending	Medicare Part A and B spending ($/beneficiary/month)	Intervention quarters 1, 2		1,802	-8.6 (69)	-0.5%	0.451
		Intervention quarters 3, 4		1,774	19.2 (71.2)	1.1%	0.606

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison group, divided by the adjusted comparison group mean.

[b] The *p*-values from the secondary test results were not adjusted for multiple comparisons within or across domains.

[c] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

FFS = fee-for-service.

Table II.C.6. Preliminary conclusions about the impacts of CareFirst's HCIA program on patient outcomes, by domain

		Evidence supporting conclusion		
Domain	Preliminary conclusion	Primary test result(s) that supported conclusion	Primary test result(s) plausible given secondary tests?	Primary test result(s) plausible given implementation evidence?
Quality-of-care outcomes	Indeterminate effect	• Neither of the individual tests in the domain was statistically significant or substantively important • The combined test across both outcomes in the domain was not statistically significant or substantively important	Yes	Yes
Service use	Indeterminate effect	• Same as above	Yes	Yes
Spending	Indeterminate effect	• The single test in the domain was not statistically significant or substantively important	Yes	Yes

Sources: Tables II.C.4 and II.C.5

As mentioned previously, these conclusions are preliminary because the analyses do not yet cover the full period that we will include in the final impact analysis in future reports. CareFirst continued to hire new care managers and provide care management and care transitions services to new high-risk Medicare beneficiaries after the end of the primary test period covered in this report. It is possible that, when we extend the final evaluation to include an additional four quarters of outcomes, the program will have measurable effects in one or more of the domains.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

CareFirst received HCIA funding to expand its commercial PCMH program to Medicare FFS beneficiaries in Maryland, which provides care coordination and care transitions support to high-risk beneficiaries, as well as financial incentives to PCPs. The program aims to reduce Medicare spending while improving quality. After an initial year-long delay, CareFirst implemented the HCIA-funded initiative largely as intended, aiming to mirror its commercial PCMH program. Program implementation was also facilitated by having highly engaged PCPs, integrating HCIA-funded LCCs into care teams, and using HCIA-funded program consultants to share data with PCPs on beneficiaries' patterns of care and service use. Implementation was hindered by challenges in identifying who would benefit most from care plans due to lags in claims data used to identify high-risk beneficiaries, as well as the overall complexity of addressing the medical and social needs of Medicare beneficiaries compared with commercial patients. The HCIA-Primary Care Redesign Clinician Survey found that most clinicians believed

the HCIA-funded initiative would have a positive effect on quality of care, patient-centeredness, and clinicians' ability to respond in a timely way to patient needs.

The impact evaluation found no measurable effects of the program on quality-of-care outcomes (30-day readmissions or hospitalizations for ACSCs), service use (all-cause hospitalizations or outpatient ED visits), or Medicare Part A and B spending for high-risk Medicare beneficiaries during the first six months of the primary test period (months 12 through 18 after the program began). For service use, the statistical tests were well powered to detect effects, so the lack of measured effects is likely because the program truly did not have substantively large effects. In contrast, for the other domains (quality of care and spending), the lack of measured effects might be because the program did not have effects or that it did but, due to modest statistical power, our tests failed to detect them. The program could have measurable impacts in one or more of three domains when the evaluation is extended to cover the full primary test period (months 12 through 30 after the program began).

Our next steps for this evaluation are to (1) monitor CareFirst's ongoing program implementation and any plans for sustaining the program beyond the funding period by reviewing quarterly data submitted by CareFirst, (2) evaluate trainees' and clinicians' attitudes and experiences with the program in the third year of the award through administered surveys, and (3) extend the impact evaluation to include the full period of program operations, and (4) use the implementation findings to help interpret the impact results.

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table A.1. Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

Chronic Conditions Data Warehouse. "Table B.2. Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014b. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf. Accessed November 19, 2014.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Gilman, Boyd, Sheila Hoag, Lorenzo Moreno, Greg Peterson, Linda Barterian, Laura Blue, Kristin Geonnotti, Tricia Higgins, Mynti Hossain, Lauren Hula, Rosalind Keith, Jennifer Lyons, Brenda Natzke, Brenna Rabel, Rumin Sarwar, Rachel Shapiro, Victoria Peebles, Cara Stepanczuk, KeriAnn Wells, and Joseph Zickafoose. "Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs. First Annual Report, Volumes I and II." Princeton, NJ: Mathematica Policy Research, November 14, 2014.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21.

Truven Health Analytics. *AHRQ Quality Indicators, Prevention Quality Indicators v5.0 Benchmark Data Tables*. Prepared for the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. Santa Barbara, CA: Truven Health Analytics, March 2015. Available at http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V50/Version_50_Benchmark_Tables_PQI.pdf. Accessed August 18, 2015.

U.S. Census Bureau. "2008–2012 American Community Survey, Median household income." Washington, DC: U.S. Census Bureau, 2012.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Cooper University Hospital and the Camden Coalition of Healthcare Providers

March 2016

Kristin Geonnotti
David R. Mann
Cara Stepanczuk
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850
Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Lorenzo Moreno

Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

COOPER UNIVERSITY HOSPITAL AND THE CAMDEN COALITION OF HEALTHCARE PROVIDERS

This individual program report provides a summary of the findings to date from our evaluation of the primary care redesign (PCR) program implemented by Cooper University Hospital and the Camden Coalition of Healthcare Providers (CUH/CCHP) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the CUH/CCHP program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on patients' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF CUH/CCHP

CUH/CCHP received three-year, $2.8 million HCIA funding from CMMI to expand the care coordination program it operates in Camden, New Jersey (Table I.1 summarizes key details on the award). This community-based program targets people with high rates of inpatient utilization, using multidisciplinary care teams to help program participants stabilize their medical and social conditions. Program staff work with participants for an average of 90 days, aiming to reduce the need for costly acute care services, improve health outcomes, and meet patient-centered goals. CUH/CCHP estimates it could reduce participants' inpatient and emergency department (ED) costs by 35 percent. Combined with expected increases in primary care, specialty care, and medication expenditures resulting from improved care coordination and care management, CUH/CCHP estimates it could reduce the total health care costs of its participants by 30 percent. CUH/CCHP received a no-cost extension to continue providing program services through December 2015.

Table I.1. Summary of CUH/CCHP PCR program

Awardee's name	Cooper University Hospital and the Camden Coalition of Healthcare Providers
Award amount	$2,788,457
Implementation date	October 15, 2012
Award end date	December 31, 2015
Program description	1. Deploys multidisciplinary care teams to empower participants to better manage their medical and social conditions 2. Transitions participants to primary care and social services after about 90 days 3. Leverages health information technology (health IT) to document encounters and improve program operations
Innovation components	Care coordination, care management, and transitional care
Intervention focus	Individual
Workforce development	Hired nurses, social workers, community health workers, and a behavioral health provider to form mobile care management teams; hired program managers to supervise teams and improve operations

Table I.1 *(continued)*

Target population	Medicare and Medicaid beneficiaries with chronic conditions who have high utilization of inpatient services
Program setting	Community
Market area	Local
Market location	Urban (Camden, New Jersey, a federally designated medically underserved area)
Core outcomes	• Reduce inpatient stays and ED visits, resulting in a reduction of inpatient and ED costs of 35 percent • Increase appropriate use of primary care, specialty services, and medication resulting in a reduction in total health care costs of participants by 30 percent

Source: Review of CUH/CCHP program reports, March 2015.

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how it has been adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external factors. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of CUH/CCHP's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in April 2014 and March 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

CUH/CCHP's care management program relies on multidisciplinary care teams to empower participants to better manage their conditions, aiming to reduce their need for acute care services. The program engages patients and coordinates the services required to stabilize their medical and social needs, and promotes the consistent use of preventive services and self-care. At the start of the program, the care team develops a care plan that reflects participants' self-identified goals with an emphasis on moving the participant toward independence during the intensive program.

Specifically, the care team helps participants secure appointments with primary care providers (PCPs) and specialists, coaches participants for office visits, arranges transportation for and accompanies participants to office visits, conducts home visits, offers guidance by telephone, helps link participants to social services (such as housing or Social Security benefits), and promotes self-management of chronic diseases.

b. Target populations, risk assessment, recruitment, and enrollment

Participants must be Camden-area residents from 18 to 80 years of age with two or more hospital admissions in six months, two or more chronic conditions, with health insurance, who also meet two of five program criteria that suggest they need help managing their condition (see Table II.A.1 for details). The program identifies participants using a citywide health information exchange (HIE) and inpatient electronic health record (EHR) systems. A triage specialist manually verifies eligibility through the EHRs before enrollment. Prior to receiving HCIA funding, CUH/CCHP forged data-sharing agreements with hospitals in Camden and developed an HIE that enabled CUH/CCHP to identify potential participants by viewing real-time data about patients admitted to Camden city hospitals. Enrollment specialists invite eligible patients to enroll in the program while they are still admitted to the hospital (see Table II.A.1).

Table II.A.1. Key details about program design and adaptation

	Community-based care coordination/care management
Target population	Adults (ages 18–80) with insurance living in or near Camden with two or more hospital admissions in six months and two or more chronic conditions who meet two of these five criteria: five or more outpatient medicines, difficulty accessing service, insufficient social supports, mental health comorbidity, and active user of drugs or is homeless.
Patient Identification	After using citywide HIE and inpatient EHRs to screen for eligible patients, enrollment specialists conduct a pre-enrollment visit to patients while they are still admitted. Pre-enrollment includes describing the care management program, assessing interest in participation, and obtaining consent. Hospital-based care management staff may also engage patients to build trust, discuss patients' goals, and promote safe discharge.
Patient recruitment and enrollment	Community-based care management staff complete the enrollment process at a one- to two-hour home visit, which is expected to be scheduled within 72 hours of discharge. The initial home visit includes a full medication reconciliation, psychosocial assessment, and development of a care plan based on the participant's goals.
Service delivery protocol	Community-based care teams develop care plans to help participants better manage their conditions. Care teams see participants once a week for at least 90 days (which can be extended if circumstances such as substance abuse, homelessness, or behavioral health issues warrant), at which point they are transitioned to primary care and social services.
Adaptations	Toward the end of program implementation: • The HIE was expanded to also include information from select Trenton-area hospitals. • A hospital-based care team began engaging patients who were still in the hospital (during pre-enrollment). • A behavioral health consultant began to train care team staff and address participants' behavioral health and substance use issues.

Source: Interviews from second site visit, March 2015; document review, March 2015.

Note: Participants with admissions related to oncology, acute disease, injuries, surgeries for acute conditions or injuries, and chronic conditions for which treatment is limited are excluded from the program. Note that the program *includes* participants with lengthy subacute stays. Staff continue to follow them during their stay, making periodic (ideally weekly) contact, then beginning intensive services when participants are discharged to the community.

EHR = electronic health record; HIE = health information exchange.

c. Intervention staff and workforce development

CUH/CCHP's program includes two community-based care teams, one hospital-based care team, and one social work team (see Table II.A.2 for details). Neither community-based care team works within a primary care practice setting; instead, they work with participants in their homes and community settings (for example, if a participant is homeless). Program managers and registered nurse (RN) supervisors co-manage the community-based care teams. RNs are responsible for clinical guidance and program managers focus on operational efficiency and effectiveness. CUH/CCHP began operating a hospital-based care team, consisting of two clinical staff and two health coaches, in July 2014. CUH/CCHP believes this gives the program a "head start" and helps them better track enrollees during the pre-enrollment phase.

Table II.A.2. Key details about intervention staff

	Staff responsibilities by team		
Staff position	**Hospital-based staff**	**Community-based care teams (2 teams)**	**Additional supports (work across care teams)**
Registered nurse (RN)	One RN (not HCIA-funded). The RN conducts a chart review and psychosocial assessment for enrolled participants who are still in one of two Camden hospitals. Along with the social worker, the RN focuses on learning participants' goals, identifying potential barriers to care, and setting expectations for participation in the program. They also try to arrange the follow-up PCP appointment and begin to connect participants to social services. The RN and social worker team visit participants daily to develop relationships that can help participants stay engaged with the program after discharge.	One RN supervises each team. The RN is responsible for developing individual participants' care plans, delegating duties and responsibilities among staff on her care team, and managing relationships with community-based PCPs and clinics. In addition, RNs support the development of clinical policies and procedures and ensure that regulatory guidelines and standards are met.	--
Program manager	--	One program manager per team. Program managers (who do not necessarily have a clinical background) help staff improve internal workflows and generally ensure they have administrative support for the complex patient care they provide. Program managers use performance metrics and regular meetings with their staff to identify and mitigate inefficiencies in program implementation. In addition, program managers help care teams work through challenges they encounter when providing care to participants.	One social work program manager. The program manager (who has a master's degree in Social Work) coordinates the workload of the social work team when the care teams request social work assistance. She also helps the social work team address challenges they encounter when providing care to participants.

Table II.A.2 *(continued)*

	Staff responsibilities by team		
Staff position	**Hospital-based staff**	**Community-based care teams (2 teams)**	**Additional supports (work across care teams)**
Licensed practical nurse (LPN)	--	Two LPNs per team. LPNs are responsible for the bulk of care management and care coordination activities relating to medical needs post-discharge. LPNs conduct the initial home visit (at which the participant's enrollment is completed), develop a care plan based on participants' goals, help arrange and accompany participants on medical appointments, and provide disease education and guidance on self-management, among other activities.	--
Community health worker (CHW)	--	Two CHWs per team. CHWs assist the LPNs with care management activities to meet participants' needs post-discharge. Common activities include mitigating language barriers, conducting follow-up calls and assessments, coordinating medical equipment or transportation, and building relationships with the participant and his or her family.	--
Health coach (HC)	Health coaches may assist the RN and social worker with care coordination activities to meet participants' needs during the pre-enrollment phase as needed.	Two health coaches per team. Health coaches are AmeriCorps volunteers who work with participants who are medically stable but have not yet reached social goals after the 60-day mark. HCs help participants address social conditions and prepare for graduation but alert LPNs if any medical needs emerge.	--

Table II.A.2 *(continued)*

Staff position	Staff responsibilities by team		
	Hospital-based staff	Community-based care teams (2 teams)	Additional supports (work across care teams)
Social worker	One social worker (not HCIA-funded). The social worker teams with the RN to conduct a psychosocial assessment for pre-enrolled participants who are still in one of two Camden hospitals. The team focuses on learning participants' goals, identifying potential barriers to care, and setting expectations for participation in the program. They also try to arrange the follow-up PCP appointment and begin to connect participants to social services. The team visits participants daily to develop relationships that can help participants stay engaged with the program after discharge.	--	Two social workers (one HCIA-funded). The social work team serves in a consultative role to the care teams if care teams discover complex social conditions they cannot handle, which can include mental health, substance abuse, and housing issues. One social worker specializes in handling housing issues exclusively. The other social worker handles other issues.
Enrollment specialists	Two enrollment specialists (not HCIA-funded). The enrollment specialists conduct a pre-enrollment visit to potential participants while they are still admitted. Pre-enrollment includes describing the care management program, assessing interest in participation, and obtaining consent. Enrollment specialists hand the participant off to the hospital-based care team after they obtain consent.	--	--

Table II.A.2 *(continued)*

Staff position	Staff responsibilities by team		
	Hospital-based staff	Community-based care teams (2 teams)	Additional supports (work across care teams)
Behavioral health provider	--	--	One clinical psychologist. The psychologist provides limited clinical services, referrals to other behavioral health practitioners, and behavioral health service coordination to participants. She also provides training to care teams on how to address the mental health issues they might encounter when working with the high utilizer population.

Source: Interviews from second site visit, March 2015; document review, March 2015.

Since the beginning of program implementation, CUH/CCHP administrators have offered training modules to enhance staff capacity. Administrators select topics that address the community-based nature of the program, largely based on staff input. For example, a safety and outreach training reinforced guidelines to ensure staff safety while engaging participants on visits in Camden-area homes and community spaces, such as designating a safe public space to meet if a house is deemed unsafe; using a code word to signal an unsafe situation; and establishing a rotating on-call manager in case of staff emergency. A de-escalation training focused on crisis situations and was provided through the recently added behavioral health contract.

CUH/CCHP also recently developed a philosophy of care, named COACH, and a set of staff trainings to implement the framework. COACH is intended to help care team members learn how to develop a beneficial, therapeutic relationship with participants through five concrete skills (one for each letter of the acronym). Topics include (1) connecting medical and social tasks to participants' goals; (2) observing participants' normal routines; (3) assuming a coaching style ("I do," "We do," or "You do") based on participants' current capabilities; (4) continuing discussions with participants regarding priorities reflected in their care plans; and (5) highlighting participants' progress through data visualization.

d. Service delivery protocols

After a hospitalized patient agrees to participate in the program, CUH/CCHP's hospital-based care team visits participants daily to build relationships, complete a psychosocial assessment, identify the participant's goals, and arrange for the follow-up PCP visit. If the participant does not already have a regular PCP, CUH/CCHP staff will help them find one. Staff

might also be able to initiate connection to social services, such as starting the process of getting a driver's license, while the participant is in the hospital. CUH/CCHP staff also increasingly advocate for participants to be discharged to subacute rehabilitation facilities when they believe it can help stabilize a participant better than a discharge to the community (for example, if the participant has an unstable housing situation). The hospital-based staff took over some of the work that used to be done during the first home visit, which slightly reduced the time spent at the first home visit by the community-based care team. However, the services provided in the hospital can still be characterized as "light touch" relative to what the community-based care team provides. CUH/CCHP staff still defined the beginning of the program as the "discharge to community" date because the community-based staff provide the intensive coaching, goal-setting, and care management services to participants.

When a participant has been discharged to home, the community-based care team takes over participant contact. Under the COACH philosophy, care team members prioritize the patient-centered nature of the program. One of the first activities care teams do with participants is called backward planning, during which care teams discuss with participants their primary goals and develop a care plan that follows directly from participant-identified goals (even if these would not have been the goals that the care team member would have identified or prioritized for the participant). During the course of the program, care team members engage participants at least weekly (but often more frequently) to deliver a variety of care management and care coordination services, which can be categorized into 15 domains under two broad categories (health and social services), as shown in Table II.A.3:

Table II.A.3. Domains of care management and care coordination services

Health	Social services
Addiction	Advocacy and activism
Health maintenance, management, and promotion	Benefits and entitlements
Medication and medical supplies	Education and employment connection
Mental health support	Family, personal, and peer support
	Food and nutrition support
	Housing and environment
	Identification support
	Legal assistance
	Provider relationship-building
	Patient-specific (wildcard)
	Transportation support

Source: Interviews from second site visit, March 2015.

Care teams introduce the idea of graduating from the program within 90 days at pre-enrollment and engage participants on progress toward goals and potential for graduation throughout the program. The decision to graduate a participant is subjective; community-based care teams discuss each participant's potential for graduation during weekly care planning meetings with RN managers. Staff assess the extent to which participants have reached their goals. Participants may graduate if they have met their care plan goals, been able to minimize

unnecessary hospital and ED use, can complete important tasks on their own (such as scheduling an appointment), and can be connected to a community resource or PCP that can help them coordinate their care. Recognizing that not all goals are achievable within 90 days, staff must determine whether they can help participants progress further or if it is more appropriate to refer them to other resources in the community. Social conditions can take longer to address than medical issues; after 60 days, community-based health coaches often take primary responsibility for engaging participants who have been medically stable, but have outstanding social issues.

2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying in part on interviews with program administrators and self-reported information included in CUH/CCHP's quarterly self-monitoring and measurement reports. Table II.A.4 summarizes CUH/CCHP's self-reported program implementation measure targets and achievements.

a. Program enrollment

From its inception through March 2015, the program has enrolled 428 patients (70 percent of the three-year projection of 610 enrollees). Program participants have an average of seven chronic conditions, including hypertension, diabetes, depression, anxiety, hyperlipidemia, and asthma. The acceptance rate—calculated as the total number of eligible patients enrolled divided by the total number of eligible patients approached—decreased over time, particularly since September 2014. The acceptance rate was 59 percent in March 2015, below the program's target of 75 percent. Of those participants who exited the program (380 participants), 66 percent graduated. The reasons the remaining participants exited the program included: inability to locate the participant, inability to fully engage the participant (due to lack of receptivity to the program), and death. Program administrators cite participants' resistance to change their behavior as a major challenge to accepting program services.

b. Service measures

CUH/CCHP was largely successful in reaching its program process and service delivery goals. Administrators recognized that engaging participants quickly and consistently after discharge from the hospital increased the likelihood participants will achieve their self-identified goals. Therefore, we use the following three measures to assess the degree to which the program adhered to these principles: (1) participant encounter hours, (2) timely initial home visits, and (3) timely follow-up PCP visits.

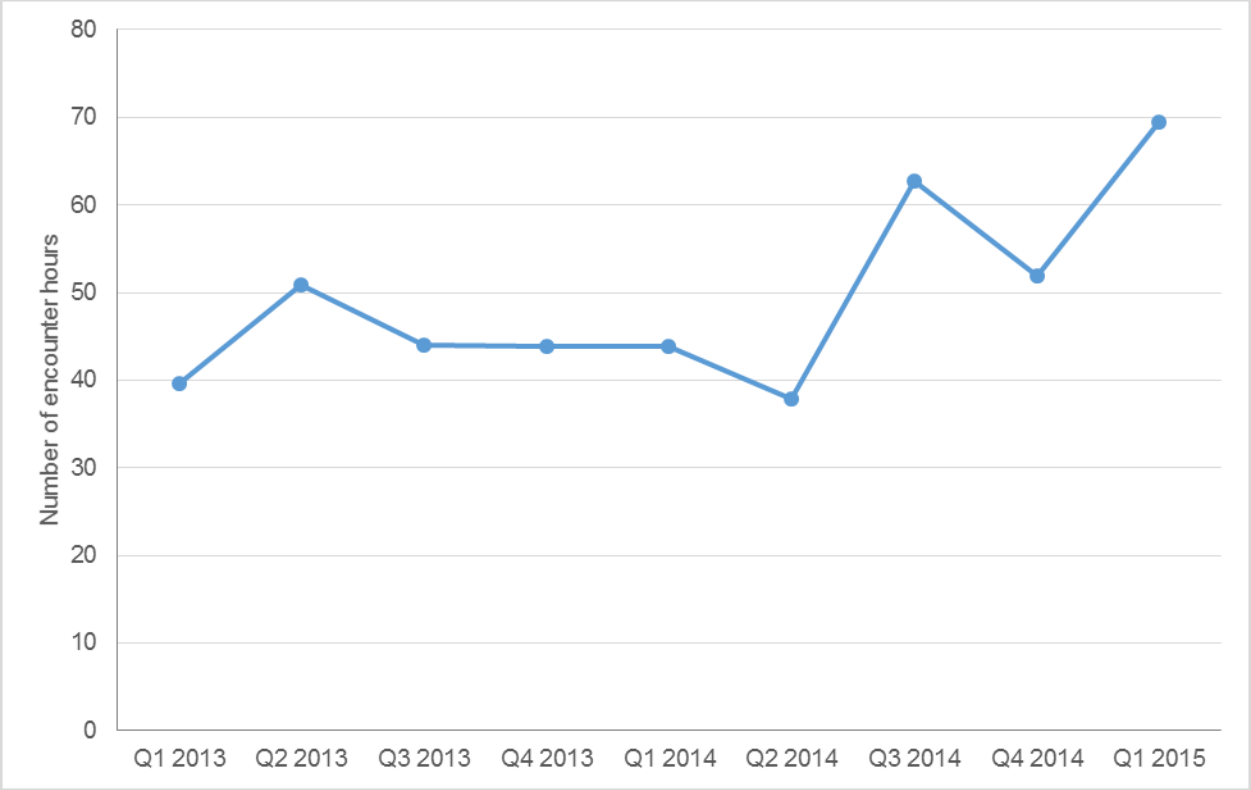
First, from program inception through March 2015, CUH/CCHP frontline staff devoted an average of 50 hours per enrollee (either with or on behalf of participants). As Figure II.A.1 shows, encounter hours per enrollee increased by 75 percent since administrators began to monitor it. The majority of time (72 percent) is spent on in-person participant encounters (community and home visits, accompaniment to medical appointments), with only a small percentage of time spent on telephone encounters and attempted contacts (4 and 5 percent, respectively).

Table II.A.4. CUH/CCHP self-reported program implementation measures

Measure	Target	Actual	Met target?	Adaptation?
Program enrollment	610	428 (through March 2015)	Did not meet in initial funding period; plan to meet during no-cost extension period	Yes: Halfway through implementation, program administrators expanded the target population to include pregnant women and two zip codes adjacent to the city of Camden.
Program acceptance rate	75 percent	59 percent (through March 2015)	No	Yes: Given the importance of participant engagement during the program, the program's lack of capacity to enroll every eligible participant, and the resources spent on persuading and retaining reluctant patients, administrators have recently instructed enrollment specialists to shift their emphasis to focus on recruiting potential participants who seem most likely to accept and participate in the program.
Patient encounter hours	Not specified	50 hours per enrollee per quarter (January through March 2015)	n.a.	CUH/CCHP uses program managers to reduce the administrative burden on care teams so that they can increase the number of hours they spend with participants.
Timely initial home visits	60 percent within 3 days	42 percent within 3 days (January through March 2015)	No	CUH/CCHP staff began to schedule initial home visits (post-discharge) with participants while they were still in the hospital. The major reasons reported for a delayed home visit are the inability to find a participant due to a weekend discharge, homelessness, a family member blocking access, or participant's lack of interest after leaving the hospital. Administrators acknowledge that this illustrates the difficulty in sometimes locating high-utilizer participants after discharge. It also highlights the importance of building relationships with participants and their families in the hospital and prescheduling the home visit.
	60 percent overall	86 percent overall (January through March 2015)	Yes	
Timely follow-up PCP visits	30 percent within 7 days	42 percent within 7 days (January through March 2015)	Yes	CUH/CCHP administrators attribute part of this success to stakeholder outreach. Specifically, they encourage local PCPs through an initiative known as the "7-Day Pledge" and through incentives related to the Camden accountable care organization structure (physicians get paid $150 for each enrollee they see within the seven-day window).
	30 percent overall	72 percent overall (January through March 2015)	Yes	

Source: Interviews from second site visit, March 2015; document review, March 2015.

Figure II.A.1. CUH/CCHP-reported average number of encounter hours per enrollee per quarter



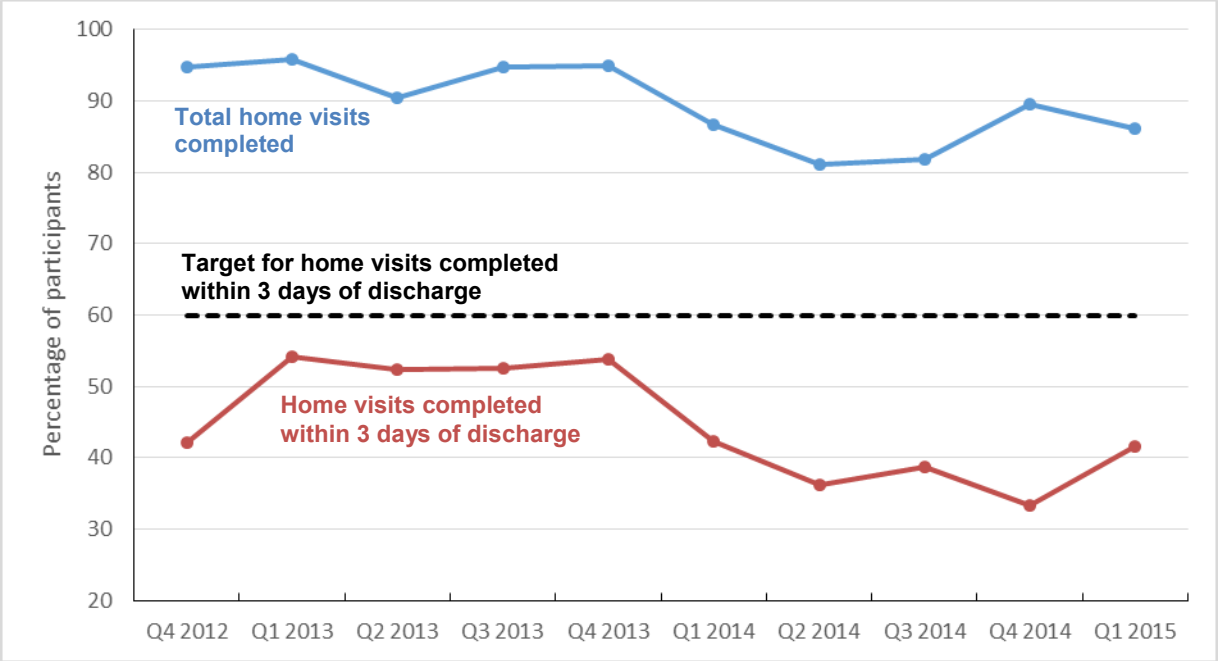
Source: Calculated from Awardee Quarterly Measures and Self-Monitoring Report, March 2015.

Note: Calculated as total number of encounter hours in the quarter divided by total number of enrolled patients in the quarter.

Q = quarter.

Second, CUH/CCHP staff have struggled to meet their goal of conducting the initial home visit within 72 hours of discharge for at least 60 percent of participants; administrators noted that most home visits happened within seven days (Figure II.A.2). From January 2014 through March 2015, the program reported that fewer than half of patients had a home visit in 72 hours.

Figure II.A.2. CUH/CCHP-reported proportion of participants per quarter with initial home visits completed and with initial home visits completed within 3 days of discharge



Source: Awardee's calculations from Awardee Quarterly Measures and Self-Monitoring Report, March 2015.

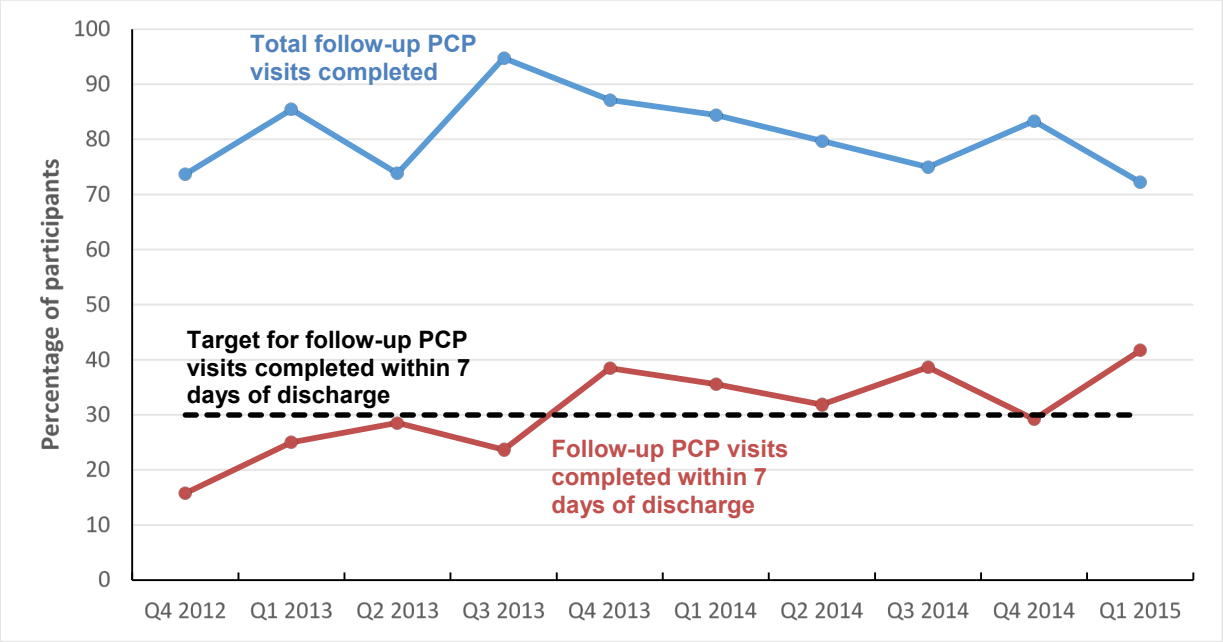
Notes: Total home visits completed were calculated as the total number of enrollees with initial home visits completed per quarter divided by the total number of enrollees per quarter.

Home visits completed within three days of discharge were calculated as the total number of enrollees with initial home visits completed within three days of discharge per quarter divided by the total number of enrollees per quarter.

Q = quarter.

Third, CUH/CCHP staff met their goal of having at least 30 percent of enrollees complete a follow-up visit with a PCP within seven days of discharge since September 2013 (Figure II.A.3). They also recorded high rates of participants with follow-up PCP visits, ranging from 72 to 95 percent of participants per quarter with a follow-up PCP visit at some time during the program.

Figure II.A.3. CUH/CCHP-reported proportion of participants per quarter with follow-up PCP visits and follow-up PCP visits within 7 days



Source: Awardee's calculations from Awardee Quarterly Measures and Self-Monitoring Report, March 2015.

Notes: Participants with follow-up PCP visits were calculated as the total number of enrollees with follow-up PCP visits completed per quarter divided by the total number of enrollees per quarter.

PCP visits within seven days were calculated as the total number of enrollees with follow-up PCP visits completed within 7 days of discharge per quarter divided by the total number of enrollees per quarter.

Q = quarter

c. Staffing measures

CUH/CCHP hired many of the key staff deemed necessary for the program by the end of its third quarter of operations (June 2013). After the first year of program operations, administrators recognized a need for more management support and the capacity to handle behavioral health issues; therefore, in its third year of operations, CUH/CCHP hired program managers and a behavioral health consultant, bringing the total of HCIA-funded staff up to 20 full-time equivalents (FTEs) (which is in line with projections). Only 9 FTEs were considered new hires, as staff from other parts of the organization were shifted to work on the program. CUH/CCHP administrators reported that they sometimes faced difficulty hiring highly qualified applicants and turnover of care team staff because of the community-based nature and emotional intensity of the work. However, CUH/CCHP reported a 94 percent staff retention rate since the program launched.

d. Program time line

CUH/CCHP implemented the program on schedule. Although the program's core components and mission have not changed, workflows and other operational details have

evolved over time as managers adapt to participant and staff needs. The program launched on time, in October 2012, and has proceeded as planned. Several adaptations have occurred in the past year. For example, in September 2014, the program hired a behavioral health consultant, a key role to address a previously unmet need for a large proportion of participants; this staff position further expanded in-house capacity to address clinical and training needs related to behavioral health and substance use issues. In July 2014, CUH/CCHP implemented a pilot program (now fully integrated into the program model) to place staff at the hospital.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.5 summarizes the major facilitators and barriers to CUH/CCHP's implementation effectiveness in each domain.

Table II.A.5. Facilitators and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Adapting the program to meet participants' and staff needs • Staff and participants' perceptions of the relative advantage of the program compared with the standard delivery of care	• Program's role as service coordinator, rather than direct-care service provider (less of a barrier over time, as CUH/CCHP proactively facilitates collaboration in community)
Implementation process	• Engaging stakeholders such as local hospitals, post-acute care facilities, PCPs, community organizations, and state political leaders • Monitoring progress to guide ongoing improvement • Engaging of program staff	• Risk of burnout among frontline staff due to intense emotional nature of the job
Internal factors	• Teamwork among program staff • Leadership commitment • Supportive environment for innovation and dissemination	
External environment	• Housing First initiative	• Intensity of participants' needs • Camden's under-resourced health care (particularly access to drug rehabilitation and mental health care) and social service infrastructure

Sources: Interviews from second site visit, March 2015; document review, March 2015.

a. Program characteristics

Two characteristics helped staff implement the CUH/CCHP HCIA program: (1) adaptation of the program to meet participants' and staff needs and (2) perceived relative advantage of the

program. CUH/CCHP's efforts to continuously adapt the program to meet participants' and staff needs is the most notable facilitator. CUH/CCHP staff have adjusted many aspects of the program design, such as adding key staff roles and redesigning participant engagement strategies. For example, CUH/CCHP added behavioral health staff to address participants' mental health needs and augment care teams' ability to effectively manage those participants. Administrators also hired hospital-based staff to increase enrollment volume, increase the likelihood of completing a timely home visit and PCP follow-up visit, and streamline the intake process. In addition, CUH/CCHP administrators augmented the care teams with mid-level administrative support to improve staff efficiency and effectiveness. Another significant change was the care team's coaching framework (COACH), which gave staff a renewed focus on prioritizing the participants' self-identified goals first, rather than coordinating on behalf of the participant. This framework helps them focus on building the participant's capabilities and work toward independence and empowerment based on what the participant identifies as most important. As one staff member noted, they used to do more for the participants but realized that if a participant didn't learn how to make his or her own appointment and arrange for transportation to get there on time, it was unlikely he or she could sustain the gains after exiting the program.

Program implementation has been facilitated by recognizing the relative advantage of CUH/CCHP's model over other care management models, particularly for a high-utilizer population that has complex medical needs. CUH/CCHP administrators believe that home visits and face-to-face interactions are more effective in addressing the needs of participants with complex conditions relative to telephonic case management programs that managed care companies often provide. As one respondent noted during the second site visit, "Telephonic case management is useless, especially for folks who don't have phones or whose minutes run out quickly, who have multiple comorbidities and bio-psycho-social needs that the person on the phone knows nothing about … You can't reconcile medications over the phone for a complex patient or show them how to take their medications correctly." This is one reason that CUH/CCHP administrators have pursued Medicaid funds originally allocated to telephonic case management services to instead be reallocated for community-based care management as one part of their sustainability plan. Staff also see a benefit to their flexible, mobile, and community-based nature. Relative to being a hospital-based staffing model, CUH/CCHP staff can assist patients in many contexts and housing situations. Most importantly, staff see an advantage to the relationship and coaching style they use with patients. Care team staff have more time than the average clinician or case manager to tailor multifaceted assistance to a participant's needs, current capabilities, and environment.

The complexity of CUH/CCHP's position as a service coordinator, rather than a direct-care service provider, has become less of a barrier to implementation over time. Program administrators still believe that their impartial role enables them to work with health systems and public agencies in Camden that might otherwise view them as competitors.

b. Implementation process

Three implementation process factors have facilitated the implementation of the CUH/CCHP program: (1) engaging other stakeholders, (2) ongoing monitoring of progress to

guide improvement, and (3) staff engagement. CUH/CCHP faces one challenging factor regarding the implementation process: the risk of burnout among frontline staff due to the intense emotional nature of their roles.

First, engaging other stakeholders has helped program staff build a robust health and social service data set to identify and learn about potential participants, coordinate safer discharges, connect participants to community services, augment community resources, and help secure the program's financial sustainability. The relationships that CUH/CCHP staff have built over the past decade with local hospitals and post-acute care facilities has enabled them to incorporate the HIE and EHRs into their complex, yet efficient, patient identification and enrollment processes. CUH/CCHP staff based at the hospital can coordinate directly with the hospitals' case management and social work staff to improve the discharge process for vulnerable participants, including advocating for some to go to subacute facilities if they do not have stable housing or necessary resources for follow-up care. These CUH/CCHP hospital-based staff have been well received by the area hospitals in which they work and have integrated themselves into the daily workflows at each hospital. In addition, the program staff continue to foster professional networks with other service providers in the community—such as housing agencies and a methadone clinic—which enables participants to obtain needed resources more quickly. Although Camden-area primary care and specialist providers are not part of the HCIA program, their cooperation and collaboration is essential to implementation and participants' success. Program staff reported that their efforts to encourage PCPs to schedule participants' follow-up visits within seven days of discharge is going well. Finally, lobbying the state's political leadership has enabled CUH/CCHP to redirect a portion of Medicaid funding for telephonic case management to community-based care management. This funding stream supports CUH/CCHP's sustainability strategy, which is discussed in detail in Section II.A.4.

Second, the CUH/CCHP program collects and analyzes a broad range of self-monitoring data to help program managers identify inefficiencies in program implementation and work toward meeting program goals. CUH/CCHP staff use real-time data feeds and alerts to enhance enrollment efforts, track participants' progress, and increase coordination within and across the different staff teams. One feature that is particularly helpful for care teams is the daily status update they receive on each participant to whom they are assigned. Alerts include participants readmitted to the hospital; participants who have been in the ED within the past seven days; participants who need a follow-up home or PCP visit; and participants who are due for a 30-, 60-, or 180-day evaluation. Program managers meet with care team staff weekly to discuss these and other metrics, helping the team troubleshoot issues. Program managers also use these metrics to identify lessons learned from high performers and disseminate those best practices to the broader team. For example, they noticed that one staff member had a particularly high rate of retaining participants. They disseminated her pre-scheduling strategy (scheduling the next week's appointment at the current appointment) to the rest of the team. In addition, CUH/CCHP staff use qualitative chart reviews and interviews with all participants who have been readmitted to the hospital to determine whether the reasons for readmission seem preventable and can be addressed through the program (and if so, how should operations change to mitigate that risk).

Third, respondents cited staff engagement as a facilitator that helps care teams meet program demands and mitigate the risk of burnout for frontline staff working with a high-utilizer population. To address this issue, they prioritize finding candidates who can withstand the intense emotional nature of the job and institute workflows and company policies to support frontline staff. Administrators responded to some of the stresses during program implementation by adding administrative staff, behavioral health staff, and trainings tailored to staff needs to support program staff. The behavioral health provider, who offers clinical assessments for participants and training to staff, has alleviated a lot of the burden cited last year stemming from serving participants with mental health issues. Before adding the behavioral health provider, some care team members did not feel fully prepared to serve participants with mental health issues; new training and clinical support has normalized mental health issues and provided staff with concrete steps to deal with high-stakes situations. In addition, respondents believed that morning huddles and guidance from program managers have given them opportunities to share experiences, provide suggestions, and receive feedback in a supportive and timely manner. Finally, staff believe that recent efforts to define the workflow and expectations have helped them feel more confident and productive in their roles.

c. Internal factors

The organizational characteristics of CUH/CCHP, as well as features of the environmental context in which the organization is located, have also influenced program implementation. Three internal factors have facilitated the implementation of the CUH/CCHP program: (1) teamwork among program staff, (2) leadership commitment, and (3) implementation climate.

First, CUH/CCHP staff believe the level of teamwork has facilitated efficient workflows and enhanced decision making. Frontline staff with different specialties work in teams to address the daily obstacles faced by program participants and to examine different aspects of a participant's progress toward his or her goals. Staff noted that the focus on team-based decision making was particularly important when considering a participant for graduation because it is a subjective assessment that requires a diverse set of perspectives. Staff noted that open communication through morning huddles and their data monitoring and analysis platform have improved morale and effectiveness. Staff also noted that the benefits of specialized teams (such as the teams for social work and hospital- or community-based care management) outweighs the coordination cost that they experience as they hand participants off to subsequent teams.

Second, the commitment and responsiveness of CUH/CCHP leadership has facilitated implementation through improved staff satisfaction and operational efficiency. Staff appreciate the organization's promotion of work–life balance and a supervisory structure that focuses on supporting and problem-solving with staff. Early on, respondents reported that a lack of managerial support within the organization decreased staff satisfaction and care team effectiveness. Program administrators responded by hiring mid-level program managers and a chief operating officer to provide staff more guidance in how to optimize their daily tasks, discuss the challenges with their jobs, clarify accountability across care team roles, and help the program operate more smoothly.

Third, CUH/CCHP's supportive environment facilitates innovation and dissemination of the model. Respondents noted that the program's core strategy is to iterate quickly as new information becomes available because there is no established guide for a community-based care management program for a high-utilizer population. Program managers noted that it was less important to point out what was going wrong and more important to see how they could change the program to make it work better for staff and participants. Staff across the program supported the self-monitoring and improvement efforts. During the second site visit, one respondent observed that weekly metrics gave staff the feeling of providing something tangible to their participants. Because the work is "… hard and frustrating … data can be a place to celebrate small successes when it is [otherwise] hard to see." In addition, CUH/CCHP administrators launched a knowledge management initiative to capture, document, and synthesize protocols of care for 15 health and social service domains. This initiative is intended to standardize service delivery and increase efficiency across care team members, while also developing a resource library that could be shared externally with the broader community.

d. External environment

Features of an organization's external environment can also influence program implementation. Two external factors present challenges to program implementation: (1) the intensity of participants' needs and (2) Camden's under-resourced health care and social service infrastructure. One external factor that has the potential to facilitate program implementation during the last six months of HCIA funding is a housing initiative for high-utilizer residents in Camden County.

Participants' needs are perceived as intense, involving chronic social issues that often prevent people from self-managing illnesses and seeking appropriate treatment. These social issues include unstable housing; poor living conditions; criminal records; lack of identification; transportation; and low levels of literacy, English language proficiency, and education. Care teams also struggle to help participants overcome the effects of adverse childhood events (such as abuse or neglect) and prior negative experiences with the health care system. Respondents agreed that, although rewarding, working with this population can be emotionally draining. As one responded stated during the second site visit, "The work will have an impact on us.… We have to recognize that and manage that. Part of it is taking care of yourself; you need to be stable to engage people that are coming to you with a lot of needs." Program leadership is aware of the need to prevent staff burnout and they encourage staff to develop boundaries with participants and to practice self-care.

The general health care environment in Camden also complicates program implementation. Respondents reported that the health care infrastructure in Camden is under-resourced and health care is sometimes poorly delivered. Long waiting lists exist for services such as drug rehabilitation and mental health specialists and the medical system often tends to blame the patient for nonadherence, rather than seeking to discover the underlying reasons for a patient's actions. Care team members reported that it can be difficult to rebuild relationships between participants and providers because of previous negative experiences with an under-resourced health care infrastructure.

One countywide initiative, called Housing First, could facilitate program implementation when it is operational in summer 2015. CUH/CCHP helped launch and continues to support the initiative, which is aimed at providing affordable and stable housing that includes optional wraparound services for high-utilizers who would not normally qualify for such housing (such as those with criminal records, active substance abuse, or lack of photo ID). The Housing First philosophy considers housing a right, noting that housing should be the first step toward addressing other medical and social issues. It is likely that a portion of program participants will benefit from this initiative, which will distribute vouchers to 50 people to enter the program over the next two years (Laday 2015). The Housing First opportunity is significant because safe, stable shelter is one of the main barriers for high-utilizers (and many current program participants) to achieving and maintaining health and social goals. As an administrator noted during the second site visit, "It's important because one segment of these patients will never get better until we get [housing] fixed."

4. Sustainability and scalability

Sustainability. CUH/CCHP administrators believe a mix of foundation and state or federal government funding (in the short term) and reimbursement from public and private payers (in the long term) can support the care coordination and care management services they provide. The New Jersey Medicaid accountable care organization (ACO) demonstration began in 2015 and is authorized for three years, with the possibility of reauthorization by the state legislature. (The New Jersey Medicaid ACO demonstration did not provide CUH/CCHP with funding to continue the program; rather, it authorized the framework under which CUH/CCHP now holds contracts with Horizon New Jersey Health and UnitedHealthcare.) The Medicaid ACOs are population-based—meaning that patients are not attributed to a particular practice, but rather to a designated geographic area—and certified Medicaid ACOs can share in savings they generate in predetermined geographic areas where 5,000 or more Medicaid beneficiaries live, regardless of whether the ACO engages with the patient. To qualify for shared savings, ACOs must also meet quality benchmarks, such as screening targets and avoiding potentially preventable hospital admissions (Cantor et al. 2014). The demonstration project certification also allows exemption from antitrust laws so that hospitals, ambulatory care providers, and community-based organizations (such as CCHP) can collectively bargain for better patient outcomes and lower costs in a given geographic region. In July 2014, CUH/CCHP administrators submitted an application to become certified as a citywide Medicaid ACO, which created a mechanism for provider coalitions to share savings through care improvement initiatives. In July 2015, they were certified to form a Medicaid ACO, which enabled them to work in conjunction with their community partners to seek long-term contracts and additional funding to develop and test new iterations of the CUH/CCHP model.

CUH/CCHP had an existing agreement with one Medicaid managed care organization and was entering into an additional contract before the New Jersey Medicaid ACO demonstration started. Both of those agreements are part of a path toward sustainability as CUH/CCHP continues to develop a Medicaid ACO. CUH/CCHP just completed the first program year of its ACO agreement with UnitedHealthcare. Based on the results of the first year of this program, CUH/CCHP administrators are exploring opportunities to increase up-front care management

fees paid by UnitedHealthcare into the program beginning on December 1, 2015, to further ensure the program's sustainability. CUH/CCHP also recently put an agreement into place with Horizon New Jersey Health, the largest Medicaid managed care organization in the state. Based on projections from the results of the UnitedHealthcare contract, CUH/CCHP anticipates that the shared savings from this contract will be sufficient to sustain the program.

CUH/CCHP administrators also continue to work with policymakers to foster an agenda that supports community-based care management initiatives, particularly for the high-utilizer population. Building in part upon early program successes, the January 2014 State of New Jersey Department of Human Services Division of Medical Assistance and Health Services contract (Section 4.6.5.K) required each Medicaid managed care organization to develop and implement a community-based care management program for a subset of members who do not receive managed long term services and supports and are either medically and socially complex or frequent users of Medicaid services. Briefly, they were required to reallocate 10 percent of resources previously dedicated for telephonic case management services to instead be used for community-based care management, which aims to provide home visits and direct coordination; as such, CUH/CCHP is a logical service provider for the managed care organizations and has already established relationships to fulfill this obligation.

Additionally, program administrators are in preliminary discussions with CUH board members about a business model for placing care coordinators in the hospital to perform discharge planning and enhanced care transitions for high-cost patients with complex conditions. This model stems from the work they have conducted with participants in the hospital before discharge, and seems to be a natural extension to a wider set of participants who could also benefit from program services by piloting several approaches to how they might offer these services in the future.

Lastly, CUH/CCHP is currently conducting a randomized controlled trial in collaboration with the Abdul Latif Jameel Poverty Action Lab to test the effectiveness of the program. CUH/CCHP administrators partially view the engagement in this evaluation as one part of a broader sustainability strategy, as positive proof of efficacy could empower them to seek sustainable funding sources from a wide array of funders.

Scalability. As they have done throughout the length of the program, CUH/CCHP administrators continue to disseminate program materials in several ways, promoting cross-site learning with the implementation sites under the Center for State Health Policy at Rutgers University (another PCR awardee) and other community-based care management/coordination organizations. CUH/CCHP administrators prioritize open-source resources whenever possible, so that community partners and participants can also easily access and adapt resources. For example, when choosing an online platform to host transforming the existing care planning resources into an open-source online library of resources, they chose an open-source online platform with maximum functionality for participants and the community. Every other month, CUH/CCHP staff also host an in-person open house for people from around the country to learn more about how the model is being implemented. They also share program materials—such as

job descriptions, evaluation materials, team structure, and enrollment forms—on their website so other programs can use or adapt their tools.

B. Description of clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at CUH/CCHP provided important insights into the implementation process. In order to provide insight from clinicians external to the HCIA program on how they interact with the program and on contextual factors that might affect implementation effectiveness, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in fall 2014, the third year of the HCIA-funded program. It is important to note that the surveyed clinicians are not funded by, and do not directly interact with, the HCIA program. Rather, they are community-based clinicians in the Camden area who provide care to at least five HCIA program participants (according to CUH/CCHP) and might interact with HCIA care team members. Given the close network of providers in the Camden community and the community-based nature of the CUH/CCHP program, clinicians' views can offer important insights on factors that might ultimately affect program implementation and participants' interactions with the HCIA-funded program.

In this section, we report on clinicians' views of their daily work and practice. First, we focus on the contextual factors that can affect their ability to support the CUH/CCHP program, including the characteristics of the practice location, career satisfaction and burnout, and barriers to providing high quality and patient-centered care. We then present data on the alignment of clinicians' views and experiences with the overall goals of the HCIA-funded innovation, as well as their awareness of CUH/CCHP's program and their perceptions of program effectiveness.

In the case of CUH/CCHP's community-based care management initiative, factors affecting clinicians and their practices might indirectly affect the success of the program. These factors include resources available in the local health care system, such as health information technology (health IT), and clinicians' attitudes toward their work environment and care management initiatives. Another important consideration is the barriers clinicians face when providing care to participants with multiple medical and social comorbidities (CUH/CCHP's target population). In short, highly functional practice locations with satisfied clinicians and sufficient resources, within a coordinated and high-capacity health system, might find it easier to support the implementation and goals of the CUH/CCHP program than those that are less functional and have fewer resources.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

CUH/CCHP had a total of 34 clinicians in the sample, 20 of whom completed responses (a response rate of 64 percent).[1] Of these 20 respondents, 15 were physicians and 5 were nurse practitioners.[2] One-third of these clinicians were practicing at a solo or two-clinician practice; the rest were evenly distributed among group practices with three or more physicians, a group or staff model health maintenance organization, a federally qualified or other community health center, a hospital run by a private for-profit or not-for-profit organization, and a medical school or university. Most clinicians in the CUH/CCHP sample (75 percent) reported a fixed salary as their primary source of compensation.

Clinicians in the CUH/CCHP sample reported working in settings that are considered advanced in terms of the use of health IT. Although nationally, slightly more than one-half of physicians practice in settings with functional electronic health records (Furukawa et al. 2014), most clinicians in the CUH/CCHP sample reported using health IT at their practice locations. As shown in Table II.B.1, most clinicians reported using electronic systems for prescribing medications, entering clinical notes, drug dosing and interaction alerts, ordering tests and procedures, and accessing laboratory results. In addition, more than half of the clinicians reported using electronic systems for tracking referrals and using and computerized participant registries, functions that are generally advanced and not in widespread use nationally (DesRoches, Painter, and Jha 2014). Clinicians in the CUH/CCHP sample were unlikely to offer patient-facing technologies, with few clinicians offering their participants the option to do the following online: request a prescription refill, email a clinician about a medical question or concern, and request an appointment.

Table II.B.1. Electronic functionalities at practices

Survey item	Number of respondents	Percentage of respondents
Percentage reporting using each of the following at least occasionally		
Entry of clinical notes	19	95%
Prescribing	19	95%
Alerts warning of drug dosing or drug interactions	17	85%
Ordering of tests and procedures	16	80%
Access to laboratory test results	15	75%
Referral tracking	13	65%
Participant lists or registries	12	60%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

[1] Calculated according to American Association for Public Opinion Research standards. The response rate of 64 percent represents the number of completed responses (N = 20) divided by the total sample (N = 34) minus ineligible responses (N = 2).

[2] The number of clinicians in each response category (here and throughout this section) do not always sum to the total number of CUH/CCHP respondents (N = 34) due to survey item nonresponse, as well as clinicians who reported that a given question did not apply to their practice and thus did not provide a response.

b. How clinicians experience their careers and workdays

Clinicians practicing in the Camden community are important external partners for CUH/CCHP's care management program; therefore, it is important to examine their career satisfaction, level of burnout, and perceptions of their practice environments. These factors can all have an effect on the success of program implementation and participants' outcomes. Clinicians in the CUH/CCHP sample are generally satisfied with their careers in medicine: 80 percent of respondents were somewhat or very satisfied with their careers. However, almost half of respondents reported experiencing at least occasional burnout. Some reported experiencing one or more symptoms of burnout (including physical and emotional exhaustion), though only a few reported that burnout symptoms were persistent.

Fewer than half of clinicians in the CUH/CCHP sample felt that the amount of work they were expected to complete each week was reasonable. Slightly more than half of the clinicians in the CUH/CCHP sample also reported spending a significant amount of time each week (25 to 74 percent of their time) on tasks that could be performed by someone with less training.

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. Clinicians were almost evenly split on whether it was possible to provide high quality care to all participants: 43 percent of respondents strongly or somewhat agreed with the statement "It is possible to provide high quality care to all of my patients," whereas 53 percent strongly or somewhat disagreed. Almost all responding clinicians (89 to 95 percent) cited six barriers to providing optimal care: (1) lack of time to spend with participants, (2) lack of timely information about care provided to participants by other physicians, (3) difficulties obtaining specialist referrals, (4) participants' inability to pay for care, (5) difficulties obtaining specialized diagnostic tests or treatments, and (6) insufficient reimbursement.

c. Clinicians' perceptions of care team functioning

More than two-thirds (68 percent) of clinicians in the CUH/CCHP sample reported working as part of a care team within their practice, though it is important to note that these are likely not the CUH/CCHP care teams funded by HCIA (which are community- rather than practice-based). Therefore, these survey responses might provide some context as to the operational capacity of Camden-area practices, but do not reflect the ways in which these clinicians interact with the CUH/CCHP care teams. Overall, their perceptions of how their own practice-based care teams function was positive. Most clinicians (87 percent) agreed that members of the practice-based care team had sufficient time for participants to ask questions, used common terminology when communicating with one another (87 percent), relayed information in a timely manner (80 percent), and verbally verified information they received from one another (67 percent). Few agreed that practice-based care team staff followed a standardized method of sharing information when handing off patients.

d. Alignment with goals of PCR

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The

inclusion of the extremely important category helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are merely important. In Table II.B.2, we present results based on the proportion of clinicians rating each of these goals as extremely important. The views of clinicians in the CUH/CCHP sample aligned with the overall goals of PCR, as most clinicians rated 8 of the 13 goals as extremely important. Most notably, 75 to 85 percent of clinicians in the CUH/CCHP sample rated four PCR goals as extremely important: increasing access to primary care, improving care coordination for participants with chronic conditions, reducing hospital readmissions, and reducing ED visits. More than half of clinicians in the CUH/CCHP sample reported engagement in quality improvement initiatives—including conducting at least one clinical audit of care that participants receive and collaborating on quality improvement efforts with other practices, hospitals, government agencies, or professional associations—within the past two years.

Table II.B.2. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Increasing access to primary care	17	85%
Improving care coordination for participants with chronic conditions	17	85%
Reducing hospital readmissions	15	75%
Reducing ED visits	15	75%
Improving participants' capacity to manage their own care	13	65%
Improving care continuity in primary care	13	65%
Improving appropriateness of care	12	60%
Increasing the use of evidence-based practice in clinical care	11	55%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

3. Awareness of program and perceived effects

The overall goals of CUH/CCHP's community-based care management program are to reduce inappropriate hospital and ED utilization and improve participants' ability to receive reliable and coordinated primary and specialty care. Although the HCIA program focuses on removing barriers from the participants' perspective and improving self-management, program administrators believe that improving the capacity of local practices to help participants manage their care is also critical to long-term success. For example, if clinicians are aware of the program and believe that it will enhance the care they provide, they are likely to feel more invested in providing quality care for high-risk patients targeted by the program.

a. Perceived effect of program on participants' care

We asked clinicians who were at least somewhat familiar with the CUH/CCHP program whether they perceived favorable effects of the program on the care they provide to participants. Most clinicians in the CUH/CCHP sample (82 percent) were at least somewhat familiar with CUH/CCHP's HCIA program. Clinician's perceptions of the effect of the program on the care

they provide to participants were generally favorable.[3] More than 70 percent of the clinicians who were familiar with the program believed it would have a positive effect on the quality and equity of care for all participants, as well as on their ability to respond to participants' needs in a timely way. Most clinicians familiar with the CUH/CCHP program also believed the program would have a positive effect on patient-centeredness of care (61 percent). However, some of the clinicians familiar with the program perceived no impact within these dimensions or believed it was too soon to tell; only a few of these clinicians perceived a negative impact within these dimensions.

4. Conclusions about clinicians' attitudes and behavior

The HCIA Primary Care Redesign Clinician Survey found that most respondents were familiar with the CUH/CCHP program, even though none of the responding clinicians were HCIA-funded or members of the CUH/CCHP program's care teams. The majority of responding clinicians who were familiar with the program believed that it was having a positive effect on patient care. The majority of responding clinicians had access to advanced health IT, were satisfied with their careers in medicine, were comfortable working within a care team, and were supportive of primary care redesign goals. In addition, clinicians reported that a lack of time and information for each patient made it difficult to provide high quality care, suggesting providers may benefit from the external care management assistance provided by the CUH/CCHP program for participants with complex medical and social needs.

C. Impacts on patients' outcomes

1. Introduction

In this part of the report, we describe the design we will eventually use to estimate the impacts of CUH/CCHP's program on patients' outcomes in three domains: quality-of-care outcomes, service use, and spending. We also present preliminary results. However, because the sample sizes in this report are small, we present descriptive statistics only and do not estimate program impacts.

We first describe our overall design for estimating impacts (Section II.C.2) and then the methods we used to generate the descriptive statistics presented in this report (Section II.C.3). Next, we describe the characteristics of the current treatment and control groups at baseline (Section II.C.4) and mean outcomes for the two groups in the first three months after program enrollment (Section II.C.5). Finally, we describe our next steps for the impact analyses (Section II.C.6), including specifying the primary tests that will be central to future impact analyses.

Although CUH/CCHP's program serves Medicaid beneficiaries and Medicare beneficiaries enrolled in managed care plans as well as Medicare fee-for-service (FFS) beneficiaries, due to limitations in available data we have analyzed outcomes only for the Medicare FFS population

[3] Clinicians were asked about the perceived effect of the CUH/CCHP program and the barriers to and facilitators of implementation only if they reported being at least somewhat familiar with the program.

(including those who are dually eligible for Medicare and Medicaid). Results might not be generalizable to the full population that CUH/CCHP's program serves.

2. Design for estimating impacts

Our design for estimating impacts on patients' outcomes builds on a randomized controlled trial (RCT) that CUH/CCHP is conducting with assistance from the Abdul Latif Jameel Poverty Action Lab (J-PAL); the design and implementation of this study is independent of Mathematica's evaluation. In the RCT, CUH/CCHP identifies prospective enrollees who meet the program eligibility criteria and then asks them to volunteer to receive program services. As summarized in Table II.A.1, CUH/CCHP targets adults (ages 18–80) with insurance living in or near Camden with two or more hospital admissions in six months and two or more chronic conditions who meet two of these five criteria: five or more outpatient medicines, difficulty accessing service, insufficient social supports, mental health comorbidity, and active user of drugs or is homeless. CUH/CCHP then randomly assigns those who consent to participate to either a treatment group that receives program services or a control group that does not. CUH/CCHP began enrolling eligible and consenting volunteers into the RCT in March 2014. Given CMMI's approval, we will use the RCT to define our treatment and control groups. Specifically, our treatment group will include Medicare fee-for-service (FFS) beneficiaries (and, if sufficient data become available, Medicaid beneficiaries) who CUH/CCHP randomly assigned to the treatment group. The control group will include Medicare FFS beneficiaries (and, possibly, Medicaid beneficiaries) who CUH/CCHP randomly assigned to the control group.

We will estimate program impacts as the differences in outcomes, measured in administrative claims data, for beneficiaries in the treatment and control groups in the first six months after they enroll into the intervention. We selected six months as the follow-up period because the intervention is relatively short (typically 90 days), and the awardee expects impacts to either persist or decline (but not strengthen) after the service delivery period. We will estimate impacts using multivariate regressions that adjust for a beneficiary's characteristics at enrollment. These regressions will increase the precision of the estimates and adjust for any chance differences between the groups.

As independent evaluators, we are taking two steps to confirm that CUH/CCHP's process for assigning beneficiaries to treatment and control groups is truly random, which is essential for ensuring the two groups are balanced for impact estimation. The first step, which we have already completed, is to interview the CUH/CCHP staff who conduct the randomization and confirm that the process includes safeguards for maintaining the integrity of the randomization. We learned that the CUH/CCHP staff are using appropriate randomization safeguards, including a randomization process developed by J-PAL, software that requires enrollment specialists to obtain participants' consent before revealing treatment or control group assignment, and a process to prevent those in the control group from enrolling in the program until at least 12 months after the initial randomization date (at which point they are directly enrolled in the program if they meet eligibility requirements). Second, we will compare observable characteristics during the baseline period—which we define as the 12-month period before the discharge into the community date—between the treatment and control groups to ensure that there are no more differences than would be expected by chance alone.

We had originally intended to conduct a separate impact analysis for beneficiaries who enrolled in CUH/CCHP's program before the RCT began. This analysis would have defined the treatment group as program enrollees and the comparison group as those who met eligibility criteria but who CUH/CCHP did not approach for enrollment. However, the sample sizes—particularly for the comparison group—were too small to estimate robust impacts. In consultation with CMMI, we decided not to pursue this analysis; instead, we will focus on the RCT analysis.

3. Methods for descriptive statistics presented in this report

a. Treatment and control group definitions

For this report, the treatment and control groups include Medicare FFS beneficiaries who enrolled in the RCT between the RCT's start in March 2014 and September 2014. The September 2014 cutoff date for enrollment ensured that we could potentially follow up with all sample members for at least three months before December 2014—the end of the claims period available for this report.

In addition to the RCT enrollment date criteria, the treatment and control group members included in this report met two other unique claims-based criteria. First, for each sample member, we identified the Medicare claim for the qualifying stay—that is, a patient's second inpatient stay within six months. We used patients' enrollment dates provided to us by the CUH/CCHP to help find the qualifying stay claims. Second, sample members had to be discharged into the community, which marks the beginning of the service delivery period, by September 2014. Together with the enrollment date requirement, the community discharge date requirement helps ensure we can observe sample members' claims for at least three months. To identify when each sample member was first discharged into the community after enrollment, we examined facility discharges codes in Medicare claims during or after the qualifying stay.

After applying these awardee-specific restrictions (as well as other claims-based sample restrictions applied to all awardees), the treatment and control groups consisted of 21 and 17 beneficiaries, respectively. Although the RCT's treatment:control assignment ratio was meant to be 1:1, from March 24 through June 1, 2014 patients were unintentionally more likely to be randomized into the treatment group due to an error in CUH/CCHP's randomization program, which has since been corrected. Consequently, the treatment group currently has noticeably more members than the control group. CUH/CCHP does not monitor whether participants are randomized into treatment and control groups at a 1:1 ratio, so it is possible that the experimental groups will have different sample sizes in the future, as well.

Our recent analysis of CUH/CCHP data extracts through fall 2015 suggests that the eventual final Medicare FFS analysis sample will include about 100 treatment and 100 control group members. In addition, we hope to add Medicaid beneficiaries to the analysis sample in future reports.

b. Outcome and covariate construction

We used Medicare claims from March 2012 to December 2014 for beneficiaries in the treatment and control groups to develop two types of variables: (1) **outcomes** defined for each person in their first intervention quarter (three months after discharge to the community); and (2) **covariates** that describe a beneficiary's characteristics at baseline—that is, on the day of discharge to the community. The analysis incudes only one intervention quarter (I1) because too few (fewer than 11) Medicare FFS beneficiaries enrolled in the RCT early enough to be followed up for two or more quarters within the claims period available for this report. The Centers for Medicare & Medicaid Services (CMS) does not allow reporting results based on claims data for which the sample size is fewer than 11. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. For each sample member, we calculated six outcomes during his or her first intervention quarter. We grouped these outcomes into three domains:

1. Domain: Quality-of-care outcomes
    a. Inpatient admissions for ambulatory care-sensitive conditions (ACSCs) (number/quarter); also called potentially preventable admissions
    b. Number of inpatient admissions followed by an unplanned readmission within 30 days (number/quarter)
2. Domain: Service use
    a. All-cause inpatient admissions (number/quarter)
    b. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
    a. Total Medicare Part A and B spending ($/month)
    b. Medicare inpatient spending ($/month)

Four of these outcomes—all but ACSCs and Medicare inpatient spending—are outcomes that CMMI has specified as core for the evaluations of all HCIA programs. Our definition of the readmission measure, however, differs from CMMI's standard definition. CMMI typically defines readmissions as the proportion of inpatient admissions that end in an unplanned readmission. Instead, we analyze impacts on the number of these unplanned readmissions per thousand beneficiaries per quarter because this enables us to look at the total impact on readmissions across the treatment group, rather than readmissions contingent on an inpatient admission (because the intervention might affect the number and types of admissions as well). Though CUH/CCHP did not specify its key outcome measures in great detail, it expected the program to decrease participants' net health care expenditures primarily through reducing inpatient and ED admissions.

Covariates. The covariates include (1) number of hospital and ED discharges and Medicare Part A and B spending in the six months before the intervention start date; (2) the number of major chronic conditions (among 25 mostly physical health conditions) a beneficiary had at baseline; (3) Hierarchical Condition Category (HCC) scores (CMS's risk score, which reflects a beneficiary's predicted Medicare spending in the following year); (4) whether a beneficiary is younger than 65 years, which accounts for age and proxies for reason for entitlement (those younger than 65 qualify for Medicare due to disability whereas those 65 or older qualify due to age, although they might originally have qualified due to a disability); and (5) other demographics (gender and race). These covariates will help to measure balance between the treatment and control groups and eventually be used in the impact analysis to reduce the variance of the impact estimates.

4. Baseline characteristics

In this section, we describe baseline characteristics for the treatment and control groups. Because the analysis sample will grow over time, the baseline characteristics for these two groups might change substantively in the future relative to the statistics presented in this report.

a. Treatment group

At the start of their intervention periods, about half (48 percent) of the 21 treatment group members were younger than 65 (Table II.C.1). Because non-disability Medicare eligibility does not begin until age 65, the age categorization also revealed that at least half of the RCT treatment group was originally eligible for Medicare because they had a disabling condition. This is a much higher fraction of beneficiaries with disabilities than in the full Medicare population, among which 16 percent were non-elderly Social Security Disability Insurance beneficiaries (Kaiser Family Foundation, 2014). Similar to the national Medicare FFS average, about 52 percent of the treatment group were male. However, at 57 percent, the proportion of treatment group members who were black far exceeded the Medicare FFS national average. The HCC risk scores and chronic condition counts for the treatment group revealed that CUH/CCHP's strategy of targeting high-risk patients had the desired effect—relative to the Medicare FFS population, CUH/CCHP program enrollees had significantly more health issues on average. For instance, no treatment group member had fewer than two chronic conditions and 76 percent had six or more chronic conditions. Similarly, the treatment group's average HCC risk score was 3.8, with 1.0 being the nationwide average for Medicare FFS beneficiaries. A risk score of 3.8 implies that, on average, the treatment group beneficiaries were predicted to have Medicare spending in the year after enrollment that was 3.8 times the national average. Consistent with their significant health needs and CUH/CCHP's program eligibility criteria, the treatment group's hospitalizations, ED visits, and Medicare spending greatly exceeded national averages. Hospitalizations per patient per quarter, for example, were 1.130 for the treatment group, compared with the national average of 0.078. At $16,860 per patient per quarter, Medicare Part A and B spending for the treatment group was 6.5 times the Medicare FFS average.

b. Control group

Slightly more than three-quarters (77 percent) of the control group's 17 members were ages 18 to 64 (Table II.C.1). About 60 percent of control group members were male (relative to the

national average of 45 percent) and about 53 percent were black, which was substantively above the national average of 18 percent. Their average HCC score was 4.4, suggesting that the control group members were, on average, predicted to have Medicare spending in the year after enrollment that was 4.4 times the national average. Among control group members, 71 percent had six or more chronic conditions, whereas the rest had two to five chronic conditions. For service use in the six months before program enrollment, control group members averaged 1.4 inpatient admissions, 2.3 ED admissions, and $16,913 in Medicare Part A and B spending per person per quarter.

c. Similarity between treatment and control groups at baseline

Demonstrating that the treatment and control groups are similar at the start of the intervention is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying the model we will eventually use to estimate impacts—that the intervention period outcomes for the control group represent what would have happened for the treatment group, had the treatment group not received the intervention. Randomization ensures that there is no observable or unobservable factor other than the intervention to which the change in outcomes can be attributed. In RCTs, the two groups should be similar when sample sizes are large; however, the groups can differ substantially when samples are small.

Table II.C.1. Characteristics of the treatment and control groups at baseline for Cooper University Hospital and the Camden Coalition of Healthcare Providers

Characteristic of group	Treatment group (N=21)	Control group (N=17)	Absolute difference[a]	Standardized difference[b]	Medicare FFS national average
Demographic characteristics					
Age (years)					
18-64 (%)	47.6	76.5	-28.9	-0.601	18.5[c]
65+ (%)	52.4	23.5	28.9	0.601	81.5[c]
Male (%)	52.4	58.8	-6.4	-0.126	44.7[d]
Race: black (%)	57.1	52.9	4.2	0.082	18.2[d]
Health status and chronic conditions					
HCC risk score	3.8	4.4	-0.6	-0.363	1.0
Chronic conditions (# out of 25)[e]					
2-5 (%)	23.8	29.4	-5.6	-0.124	NA
6+ (%)	76.2	70.6	5.6	0.124	NA
Mean service use and spending 6 months before enrollment					
All-cause inpatient admissions (#/person/quarter)	1.130	1.395	-0.265	-0.320	0.074[f]
Outpatient ED visits (#/person/quarter)	0.904	2.324	-1.420	-0.391	0.105[g]
Medicare Part A and B spending ($/person/quarter)	16,860	16,913	-53	-0.004	2,581[h]

Source: Analysis of CUH/CCHP data and the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the treatment and control groups.

[b] The standardized difference is the difference in means between the treatment and control groups divided by the standard deviation of the variable, which is pooled across the treatment and control groups.

[c] Centers for Medicare & Medicaid Services (2013). Ratios calculated by summing (1) the number of Medicare FFS beneficiaries age 18-64 and (2) the number of Medicare FFS beneficiaries age 65+ and dividing each by the total number of Medicare FFS beneficiaries in 2012 (37,214).

[d] Chronic Conditions Warehouse (2014a, Table A1).

[e] We use 25 of the 27 chronic condition categories defined by the Chronic Conditions Data Warehouse (see https://www.ccwdata.org/web/guest/condition-categories). We exclude the Alzheimer's Disease and the Acute Myocardial Infarction flags because other flags include these conditions.

[f] Health Indicators Warehouse (2014).

[g] Gerhardt et al. (2014).

[h] Boards of Trustees (2013).

*/**/*** Significantly different from zero at the .10/.05/.01 level, two-tailed test. No differences were significantly different from zero at the .10 level.

ED = emergency department; HCC = Hierarchical Condition Category; NA = not available

Because the sample for this report is very small—21 treatment and 17 control group members—we were not surprised that the differences between the two groups were sometimes large (despite the randomization) (Table II.C.1). The standardized differences across only 4 of the 10 baseline characteristics are within our target of 0.25 standardized differences (the 0.25 target is an industry standard; see, for example, Institute of Education Sciences [2014]). Using this threshold, we observe that the treatment and control groups were dissimilar across the age categories but similar by gender and race. The two groups also differed by HCC risk score but were similar by chronic condition counts. Acute care admissions differed for the two groups, but spending was quite similar.

5. Unadjusted outcomes

Except for one outcome (inpatient admissions for ACSCs) all outcomes were substantively different between the treatment and control groups (Table II.C.2). Furthermore, the outcome differences were in the direction one would expect if CUH/CCHP's program had the desired impacts—relatively lower health care utilization and expenditures for the treatment group. Although the treatment and control groups have noticeably different mean outcomes during the intervention period, these differences should not be considered substantive evidence of program impacts because of the analysis sample's limited size.

6. Next steps

a. Overview

We plan to take four steps over the next year that will enable us to estimate program impacts on patients' outcomes for the next annual report. First, we will continue adding patients to the RCT analysis sample. The sample size will grow as CUH/CCHP continues enrolling patients into the RCT, the claims period available for future reports expands, and (if available) Medicaid beneficiaries are added to the analysis sample. We expect the sample size will grow to at least 100 treatment and 100 control Medicare FFS beneficiaries by the end of the 2015 calendar year. If the analysis sample size does not grow beyond 200 patients, the study will be poorly powered to detect substantively important impacts on patients' outcomes, though it will be well powered to detect the large impacts the awardee mentioned in its initial HCIA application. However, if the sample size substantively exceeds 200 patients, then the study will be more able to detect substantively important impacts. Second, we will continue monitoring the baseline characteristics of program enrollees for indications that randomization is being conducted correctly. Specifically, we will examine whether there is balance in baseline characteristics between the treatment and control groups. If there is balance, then there should be no more statistically significant differences between the experimental groups than one would expect to find by chance, given the number of comparisons. Also, the experimental groups should roughly be the same size. Third, we will use multivariate regression analysis to estimate impacts on outcomes in each of the three domains. (Appendix 2 is not included in this report because we do not use regression analysis. In future reports, it will provide details on the regression methods.) Fourth, we will conduct primary tests to draw conclusions within each outcome domain.

Table II.C.2. Sample sizes and unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and control groups for Cooper University Hospital and the Camden Coalition of Healthcare Providers, by quarter

	Intervention Quarter 1			Intervention Quarter 2		
	T	C	Diff (%)	T	C	Diff (%)
Number of Medicare FFS beneficiaries (unweighted)	21	17	n.a.	-[a]	-[a]	-[a]
Inpatient admissions for ambulatory care-sensitive conditions (#/1,000 beneficiaries/quarter)	190.5	235.3	-44.8 (-19.0%)	-[a]	-[a]	-[a]
30-day unplanned hospital readmission rate (#/1,000 beneficiaries/quarter)	190.5	764.7	-574.2 (-75.1%)	-[a]	-[a]	-[a]
All-cause inpatient admissions (#/1,000 beneficiaries/quarter)	428.6	1,117.6	-689.1 (-61.7%)	-[a]	-[a]	-[a]
Outpatient ED visit rate (#/1,000 beneficiaries/quarter)	548.6	2470.6	-1,922.0 (-77.8%)	-[a]	-[a]	-[a]
Medicare Part A and B spending ($/beneficiary/month)	$4,319	$7,179	$-2,860 (-39.8%)	-[a]	-[a]	-[a]
Medicare FFS inpatient spending ($/person/month)	$1,832	$5,006	$-3,174 (-63.4%)	-[a]	-[a]	-[a]

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services. See Table II.C.1 for sources for the Medicare FFS averages.

Note: All beneficiaries in the sample were enrolled under the RCT on or after March 24, 2014. The intervention quarters are measured relative to each beneficiary's date of discharge into the community. For example, the first intervention quarter (I1) for a beneficiary discharged to the community on March 24, 2014 would run from March 24, 2014 to June 23, 2014. In each intervention quarter, the sample includes Medicare FFS beneficiaries who enrolled early enough to be potentially followed up for all 91 or 92 days in the quarter and whose outcomes were observable on the date of community discharge and for at least one other day during the quarter. See text for details.
The difference between the treatment and control groups in a quarter is calculated by subtracting the mean outcome for the control group from the mean outcome for the treatment group. The percent difference equals that difference divided by the mean outcome for the control group.

[a] Not reported due to small sample size (fewer than 11 beneficiaries)

C = control group, Diff = difference, ED = emergency department, FFS = fee-for-service, I = intervention, n.a. = not applicable, RCT = randomized controlled trial; T= treatment group

A key element of the design is to estimate impacts for a limited number of specified primary tests. Because these primary tests will be central to our future methods for estimating impacts, we describe them in detail here.

b. Primary tests

Table II.C.3 shows our primary tests for CUH/CCHP, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important. The purpose of these primary tests is to focus the evaluation on

hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for detail and a description of how we selected each test). We provided both the awardee and CMMI an opportunity to comment on the primary tests.

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** As stated in Section I, CUH/CCHP aims to reduce inpatient and ED spending by an average of 35 percent for all program enrollees, increase spending for more appropriate care categories (such as primary care provider visits and prescription medications) by about 5 percent, and yield a net savings of 30 percent of current medical spending. Consequently, we will conduct primary tests examining all-cause inpatient admissions, the outpatient ED visit rate, ACSC admissions, and 30-day unplanned hospital readmissions for Medicare and Medicaid FFS beneficiaries; Medicare Part A spending and Medicare Part A and B spending for Medicare FFS beneficiaries; and Medicaid spending for Medicaid FFS beneficiaries.
- **Time period.** CUH/CCHP's impact on outcomes is most likely to emerge within about 90 days after the enrollment date—the service delivery period—and then persist or decline. Therefore, the primary tests for CUH/CCHP will examine outcomes for all Medicare or Medicaid FFS beneficiaries in the first and second quarters after they are enrolled in the program and discharged into the community. It is possible that, for the third annual report, we could have Medicaid data for CUH/CCHP covering the period through March 2015. However, *this assumption is intentionally optimistic*. Whatever the availability of Medicaid data is, we will conduct the primary tests using all beneficiaries for whom we can observe outcomes for at least one intervention quarter.
- **Population.** CUH/CCHP's program is meant to influence outcomes for all program enrollees. CUH/CCHP has not identified any program subgroup that it expects to have different program impacts from other enrollees. Therefore, the population for the primary tests will include all Medicare FFS enrolled in the program and, if data are available, all Medicaid beneficiaries enrolled in the program.
- **Substantive thresholds.** The awardee's program goals, which were stated using somewhat different outcome measures than we are using to specify our primary tests, mentioned 35 percent reductions in inpatient spending and 30 percent reductions in total spending and service use outcomes. By comparison, according to Peikes et al. (2011), patient-centered medical home (PCMH) models can plausibly expect 15 percent reductions for these outcomes among high-risk patients. CUH/CCHP serves a much higher-risk population than a typical PCMH program, which is probably why CUH/CCHP expects larger impacts than are typical among successful PCMH programs. However, we believe impacts even of the magnitude described by Peikes et al. (2011) would be meaningful to CMMI. Therefore, the thresholds for all primary test outcomes—including those for which CUH/CCHP did and did not give targets—are taken from Peikes et al. (2011).

Table II.C.3. Specification of the primary tests for Cooper University Hospital and the Camden Coalition of Healthcare Providers

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population[d]	Substantive threshold (impact as percentage of the counterfactual)[e, f]
Quality-of-care outcomes (2)	Inpatient admissions for ambulatory care-sensitive conditions (#/person/quarter)	Average over intervention quarters 1 and 2[c]	All Medicare FFS and Medicaid FFS beneficiaries in the treatment group	-15.0%
	Number of inpatient admissions followed by an unplanned readmission within 30 days (#/person/quarter)			-15.0%
Service use (2)	All-cause inpatient admissions (#/person/quarter)			-15.0%
	Outpatient ED visit rate (#/person/quarter)			-15.0%
Spending (2)	Medicare Part A and B and Medicaid FFS spending ($/person/month)			-15.0%
	Medicare and Medicaid FFS inpatient spending ($/person/month)			-15.0%

Notes: For all primary tests, the expected direction of effect is a decrease relative to the control group.

[a] We will adjust the *p*-values from the primary test results for the multiple comparisons made within each domain, but not across domains.

[b] The regression models will control for chance differences between the treatment and control groups during the baseline period when estimating program impacts.

[c] To implement the primary tests, we will take the average of the regression-adjusted estimates for intervention quarters 1 and 2.

[d] To specify the primary tests, we made assumptions about the Medicaid data that will be available by our third annual report. We believe that we could have Medicaid data through March 2015. However, this assumption is optimistic. If Medicaid data are not available, we will omit Medicaid beneficiaries from the primary test population. If Medicaid data are available only for a shorter period, we will keep in the primary test population all Medicaid beneficiaries for whom we can observe outcomes for at least one intervention quarter.

[e] The substantive threshold for all outcomes is equal to the reduction in admissions and spending that Peikes et al. (2011) indicated could be feasible among high-risk beneficiaries in a patient-centered medical home (PCMH) program. Peikes et al. (2011) discussed PCMH models, which the CUH/CCHP program is not. CUH/CCHP serves a much higher-risk population than a typical PCMH program, which is probably why CUH/CCHP expects larger impacts than is typical among successful PCMH programs. However, we believe impacts even of the magnitude described by Peikes et al. (2011) would be meaningful to CMMI.

[f] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention.

ED = emergency department; FFS = fee-for-service

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

CUH/CCHP received HCIA funding to expand its community-based care management program, which uses multidisciplinary care teams to address participants' social and medical needs, to more Camden-area residents. The program aims to reduce the utilization of acute care services; increase the appropriate use of primary care services, specialty care services, and medication; and reduce total health care costs. Using self-monitoring data to continually refine the program, engaging community stakeholders, maintaining a supportive climate for innovation within the organization, and promoting strong collaboration among program staff all helped to facilitate program implementation. The challenging needs of participants with complex conditions and Camden's under-resourced health care and social service infrastructure hindered implementation. The first round of the HCIA-Primary Care Redesign Clinician Survey found that most clinicians familiar with the HCIA-funded initiative believed it would have a positive effect on the quality and equity of care, patient-centeredness, and clinicians' ability to respond in a timely way to patients' needs.

Impact estimates are not yet available for CUH/CCHP because of limited sample size. The eventual impact evaluation will focus on Medicare beneficiaries and, if data are available, Medicaid beneficiaries who volunteered for the program and were randomly assigned to a treatment or control group. If program enrollment continues at its current pace, the impact evaluation will eventually have sufficient statistical power to detect large program effects.

Our next steps for this evaluation are to (1) monitor CUH/CCHP's ongoing program implementation reports through March 2016, and any plans for sustaining the program beyond the funding period, which ends December 2015; (2) evaluate trainees' and clinicians' attitudes and experiences with the program in the third year of the award through administered surveys; (3) expand the impact evaluation in the ways described in Section II.C.6; and (4) use the implementation findings to help explain the impact results.

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Cantor, J.C., S. Chakravarty, J. Tong, M.J. Yedidia, O. Lontok, and D. DeLia. "The New Jersey Medicaid ACO Demonstration Project: Seeking Opportunities for Better Care and Lower Costs Among Complex Low-Income Patients." *Journal of Health Politics, Policy and Law,* vol. 39, no. 6, 2014, pp. 1185–1211.

Centers for Medicare & Medicaid Services. "Table 2.2 Medicare Enrollment: Hospital Insurance and/or Supplementary Medical Insurance Programs for Total, Fee-for-Service and Managed Care Enrollees, by Demographic Characteristics as of July 1, 2012." Medicare and Medicaid Statistical Supplement, 2013 Edition. Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareMedicaidStatSupp/2013.html. Accessed August 25, 2015.

Chronic Conditions Data Warehouse. "Table A.1 Medicare Beneficiary Counts for 2003–2012." Baltimore, MD: CMS, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

Chronic Conditions Data Warehouse. "Table B.1 Medicare Beneficiary Counts for Chronic Conditions 2003 – 2012." Baltimore, MD: CMS, 2014b. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf. Accessed November 19, 2014.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no 9, 2014, pp. 1672–1679.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Kaiser Family Foundation. "Medicare at a Glance." Menlo Park, CA: Kaiser Family Foundation, August 2014. Available at https://kaiserfamilyfoundation.files.wordpress.com/2014/09/1066-17-medicare-at-a-glance.pdf. Accessed November 16, 2015.

Laday, Jason. "Camden County to Give Free Apartments to Homeless." *South Jersey Times,* February 23, 2015. Available at http://www.nj.com/camden/index.ssf/2015/02/camden_county_to_give_free_apartments_to_homeless.html. Accessed August 4, 2015.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Denver Health and Hospital Authority

March 2016

Tricia Higgins
Laura Blue
Lauren Hula
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sandi Nelson
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

DENVER HEALTH AND HOSPITAL AUTHORITY

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by Denver Health and Hospital Authority (Denver Health) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the Denver Health program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on patient outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF DENVER HEALTH

Denver Health received a three-year, $19.8 million HCIA to implement 21st Century Care, a program designed to transform the primary care delivery system to more effectively meet its patients' medical, behavioral, and social needs (Table I.1). Denver Health is an integrated safety-net system in Denver, Colorado. It is the largest provider of health care to Medicaid beneficiaries and uninsured patients in the state. Its facilities include eight Federally Qualified Health Centers (FQHCs), as well as urgent care facilities, an acute care facility with inpatient and emergency department (ED) services, and a managed care plan. Denver Health began implementing 21st Century Care in October 2012 in its eight FQHCs, three of which also participated in CMMI's FQHC Demonstration project from October 2011 to October 2014 to test a patient-centered medical home (PCMH) model. As part of the HCIA, Denver Health also created three new clinics for patients with the greatest health care needs. Through the 21st Century Care program, Denver Health hoped to (1) improve patients' health outcomes by 5.0 percent, based on an internal composite quality metric; (2) increase patients' satisfaction with between-visit care by 5.0 percent, without decreasing visit-based care satisfaction; and (3) decrease total cost of care by 2.5 percent, which reflects reductions of 0.7 percent in the first year of the program, 3.0 percent in the second year, and 3.4 percent in the third year on a per-person-per-year basis, relative to an inflation-adjusted baseline. Although Denver Health estimated these effects across the entire intervention population, program administrators expected to achieve the most significant cost reductions among its highest-risk patients through decreased use of expensive services, such as inpatient and ED care. Denver Health's HCIA award ended in June 2015.

Table I.1. Summary of Denver Health and Hospital Authority's 21st Century Care program

Awardee's name	Denver Health and Hospital Authority
Award amount	$19,789,999
Implementation date	October 29, 2012
Award end date	June 30, 2015
Program description	1. Stratify patients based on risk to more efficiently allocate additional resources 2. Redesign Denver Health's primary care delivery teams 3. Leverage health information technology (IT) to provide between-visit support 4. Create high-risk clinics to provide individualized care to patients with complex care needs

Table I.1 *(continued)*

Innovation components	Care coordination, care management, patient navigation, care transitions, patient-centered care. risk stratification, workflow or process redesign, population health approach to primary care, and health IT
Intervention focus	Practice
Workforce development	Added new staffing positions to expand the capacity of Denver Health's community health centers and to create three new high-risk clinics for patients with complex care needs
Target population	Primary care users with a focus on adult and pediatric patients with chronic conditions and/or behavioral health disorders who are frequent users of services
Program setting	Provider (community health centers and high-risk clinics)
Market area	Local (Denver, Colorado)
Market location	Urban
Outcomes	• 5.0 percent improvement in patients' health outcomes • 5.0 percent increase in patients' satisfaction with between-visit care • 2.5 percent decrease in the total cost of care

Source: Review of Denver Health program reports, March 2015.

Notes: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health IT systems, and undertaking other operational activities.

II. SUMMARY OF FINDINGS

In this chapter, we summarize the methodology and present the main findings of the evaluation as they relate to (1) program implementation, (2) clinicians' attitudes and behavior, and (3) patient outcomes.

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how it was adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external factors. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of Denver Health's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in May 2014 and May 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

The 21st Century Care program included three key intervention components (see Table II.A.1 for detail):

1. **Leveraged health information technology (IT) to provide between-visit support.** Denver Health invested in health IT to send patients five types of text messages: (1) appointment

reminders, (2) flu vaccine reminders, (3) well-child check-up reminders, (4) diet support messages to encourage healthy eating behaviors, and (5) tobacco cessation support. Denver Health sent the text messages using an automated system designed with HCIA funding. Patients from the target population in all risk groups (discussed in Section II.2.b) were eligible for services provided through this component.

2. **Created high-risk clinics to provide individualized care to patients with complex care needs.** Denver Health created three high-risk clinics, each with a different care model and target population. The clinic for children with special health care needs (CSHCN) worked with children with multiple chronic needs. Denver Health designed the clinic to provide patients with access to a multidisciplinary clinical team (including a physician, nurse practitioners, a physical therapist, a nutritionist, behavioral health specialists, and a social worker) during each office visit. The CSHCN model varied from usual care in that patients and their families could access multiple health and social services in one visit. CSHCN appointments were much lengthier than a typical visit—up to several hours—depending on each family's and patient's needs. The intensive outpatient clinic (IOC) was a primary care clinic that focused on high-risk adults with a primary physical diagnosis and multiple comorbidities. Compared to usual care, the IOC provided a wider range of services than a typical outpatient clinic, including dialysis. Like the CSHCN clinic, the IOC provided one-stop access to a multidisciplinary team, including physicians, nurse practitioners, an addiction counselor, behavioral health specialists, and a social worker. The third high-risk clinic, co-located at the Mental Health Center of Denver (MHCD), expanded community-based case management services to adult patients with severe mental health conditions and two or more hospitalizations in the previous year. Only patients in the highest-risk group (Tier 4, discussed further in Section II.1.b) were eligible to receive care in one of the three specialized high-risk clinics created using HCIA funds.

3. **Enhanced primary care delivery teams.** Enhanced primary care delivery teams included clinical pharmacists, registered nurses, behavioral health consultants, licensed clinical social workers, and patient navigators, who joined all eight community health centers at Denver Health. Patient navigators focused on care coordination and care transition interventions. Patients in the mid-level risk groups (Tiers 2 and 3), along with some patients in Tier 4 who were not served in the specialized clinics, were eligible to receive services from enhanced primary care delivery teams.

b. Target population and patient identification, recruitment, and enrollment

The target population for the 21st Century Care program included the following three groups:

1. All primary care patients at Denver Health (defined by the awardee as any person who had a primary care visit in the previous 18 months)
2. All patients enrolled in Denver Health's managed care plan
3. Frequent users of Denver Health services who did not fall into the previous two categories

Table II.A.1. Key details about program design and adaption

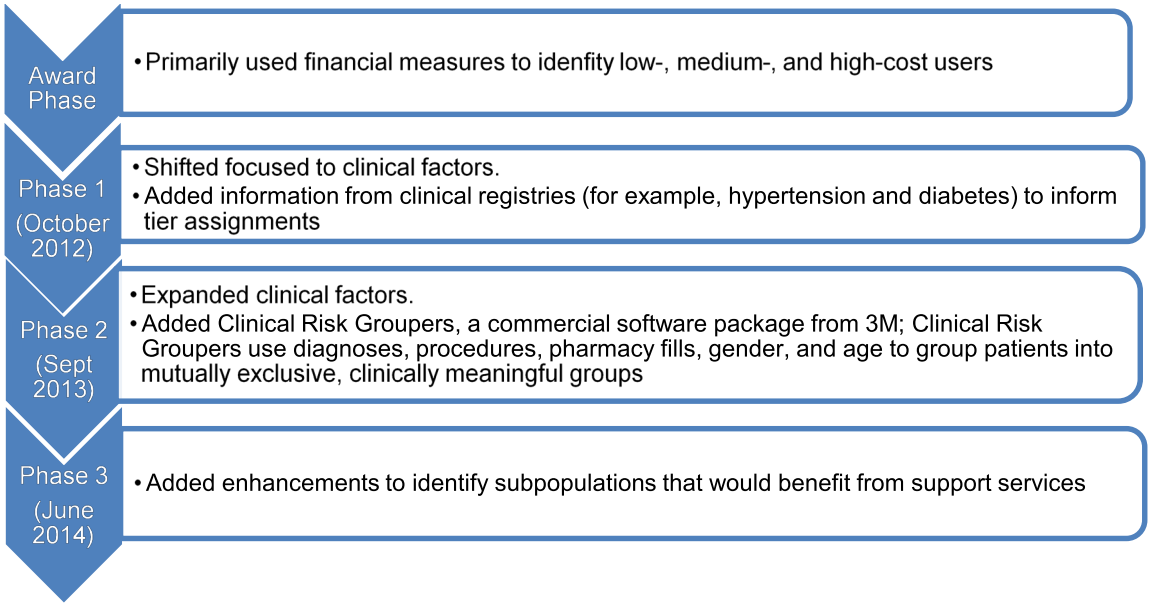
Program component	Target population	Identification strategy	Recruitment/ enrollment strategy	Intervention protocol	Adaptations
Health IT					
Automated Text Messaging	Adult, Tiers 1–4	Denver Health enrolled patients who met specific inclusion criteria. For example, adults with BMI greater than 30 who had a visit within the past six months were eligible for the diet support text message program.	Patients consented to participate in specific text messaging programs.	Denver Health sent patients text messages with appointment, flu vaccine, and well child check-up reminders; diet support to encourage healthy eating behaviors; and tobacco cessation support.	Denver Health developed a mechanism to make it easier to obtain patients' consent. Text messaging for tobacco was stopped after the pilot test period in March 2014 because participants found it confusing.
High-risk clinics					
Children with special health care needs (CSHCN) clinic	Pediatric (under 19 years of age), Tier 4	Denver Health clinically screened a subset of pediatric Tier 4 patients to determine eligibility.	Nurses triaged Tier 4 patients. If eligible for the CSHCN clinic, the patient was invited to the join the clinic. Patients' consent was required.	The CSHCN clinic worked with children with multiple chronic needs. The clinic provided a specialty consultation service designed to wrap around primary care. A multidisciplinary care team identified and addressed patients' needs.	Denver Health added a travel clinic consultation. Once a month, CSHCN clinic staff traveled to selected community health clinics to help primary care providers address the needs of their Tier 4 pediatric patients.
Intensive outpatient clinic (IOC)	Adult, Tier 4 with three hospital admissions within the past 12 months	Denver Health generated a daily list of patients eligible for care at this clinic.	Patient navigators reached out to patients recently admitted to the hospital who qualified for the IOC. Patients' consent was required.	The IOC was a primary care clinic for high-risk adults with a primary physical diagnosis and multiple comorbidities. The IOC allowed for longer visits, walk-in visits, and a higher level of care team-to-patient contact. A multidisciplinary care team identified and addressed patients' needs.	The IOC added hospital rounding in which physicians visited IOC patients in the hospital to identify barriers to care, ensure IOC participation, and help decrease the length of stay. The IOC also added home visitation for eligible patients.

Table II.A.1 *(continued)*

Program component	Target population	Identification strategy	Recruitment/ enrollment strategy	Intervention protocol	Adaptations
High-risk clinics *(continued)*					
Mental Health Center of Denver (MHCD)	Adult, Tier 4 with a severe mental health condition and two or more hospitalizations in the previous year	Denver Health identified patients.	Denver Health reached out to patients eligible for this clinic. Staff members visited eligible patients while they were in the hospital to discuss the possibility of seeking follow-up care at the clinic. Patients' consent was required.	This clinic, co-located at MHCD, expanded the clinic's existing community-based case management services to additional adult patients.	None
Enhanced primary care delivery teams					
Transitions of care	Adult, Tiers 2–4 who had been hospitalized at Denver Health	Denver Health provided this intervention to patients who received primary care at a community health center. Patient navigators identified patients using adult hospital discharge reports (run daily).	There was no formal enrollment process for this intervention. Patients could have received intervention services without knowing they were part of the HCIA program.	Patient navigators contacted patients two or three days after hospital discharge using a standardized protocol to assess transition to home needs. Navigators involved additional clinical staff as needed. Clinical pharmacists reviewed all discharged patients to identify opportunities for medication interventions.	Denver Health held several Lean events to develop and refine the transitions-of-care intervention.
Care Coordination	Adult, Tiers 3 and 4 who were assessed as high-risk and high-cost patients	Denver Health provided this intervention to patients who received primary care at a community health center. Patient navigators identified patients using high-risk care coordination reports.	There was no formal enrollment process for this intervention. Patients could have received intervention services without knowing they were part of the HCIA program.	Patient navigators contacted a list of high-risk, high-cost patients and completed adult care coordination forms. They consulted with the PCP and enhanced care team to develop a care coordination plan.	Denver Health held several Lean events to develop and refine the care coordination intervention.

Source: Interviews from second site visit, April 2015; document review, March 2015.
BMI = body mass index; CSHCN = children with special health care needs; IOC = intensive outpatient clinic; Lean = Toyota Production System's Lean methodology; MHCD = Mental Health Center of Denver; PCP = primary care provider.

Figure II.A.1. Denver Health's development of a risk-stratification tiering algorithm

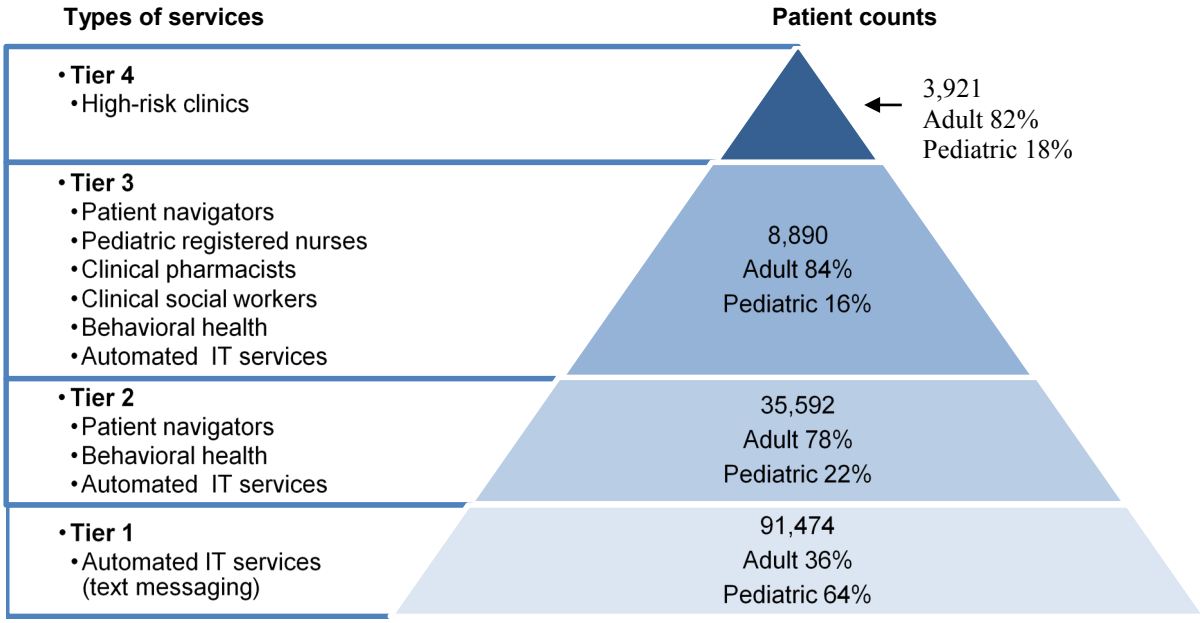


Denver Health defined frequent users as (1) people with three or more urgent-care visits, ED visits, or hospital admissions (including inpatient and observational stays) in the past 12 months; or (2) people with two or more hospital admissions, along with a serious mental health diagnosis. Qualifying mental health conditions included schizophrenic disorders and select affective and personality disorders, among others.

Denver Health used in-house administrative and clinical data to assign each patient in the target population to one of four risk-stratification tiers, with Tier 1 representing the lowest-risk patients and Tier 4 representing the highest-risk patients. As shown in Figure II.A.1, Denver Health refined this process throughout the award period. Although all iterations of the tiering algorithms preserved the financial risk-stratification goal of identifying low-, medium-, and high-cost users, Denver Health increased the clinical relevance of its tiering algorithm with each iteration. In the future, Denver Health hopes to integrate social determinants of health into its tiering algorithm.

The 21st Century Care program allocated resources across risk-stratification tiers—covering more than 100,000 patients—based on the needs of each group (Figure II.A.2). The lowest-risk tier (Tier 1) received only the new health IT component, such as text message reminders. In contrast, patients in the highest-risk tier (Tier 4) were eligible to receive care in one of the three specialized high-risk clinics created using HCIA funds. Patients in mid-level tiers (Tiers 2 and 3), along with some patients in Tier 4 who were not enrolled in the high-risk clinics, received services from the redesigned primary care delivery teams.

Figure II.A.2. Target population and types of services provided, by risk stratification tiers



Source: Denver Health and Hospital Authority, as of December 3, 2014.

Note: Denver Health stratified patients daily. As a result, the number of patients per tier fluctuated slightly each day. This figure represents the target population at a point in time.

c. Service delivery protocols

Using Toyota Production System's Lean methodology (Lean), an approach that emphasizes small-scale testing of ideas before larger implementation, Denver Health developed detailed patient intervention protocols for the transitions-of-care and care coordination components delivered by the enhanced primary care delivery teams. The Lean approach enabled Denver Health teams—including staff at multiple levels and roles in the organization affected by the intervention in question—to develop initial ideas for how to approach an intervention, pilot test it in a small number of clinics, reconvene to discuss results and refine strategies, and eventually launch the intervention on a larger scale. For example, Denver Health pilot tested the care coordination intervention to assess different strategies related to patient selection, frequency of case conference meetings, assignment of responsibilities across medical positions, and care plan documentation. Denver Health then held a series of rapid-improvement events with 21st Century Care administrative leaders, quality improvement coaches, and frontline staff to refine staff roles, improve processes, and redesign workflows. Table II.A.1 provides an overview of the finalized service delivery approaches for transitions of care and care coordination.

Over the course of implementation, each high-risk clinic developed and refined certain patient service delivery protocols, such as standardizing how patients were brought on board (or asked to join the clinics) or reminded of their scheduled appointments. However, services delivered to each high-risk patient in the CSHCN, IOC, and MHCD clinics varied significantly based the specific and complex health needs of these patients.

d. Intervention staff and workforce development

Denver Health expanded the capacity of its primary care delivery system to address patients' social and medical needs by adding new staff positions in its community health centers (Table II.A.2). As discussed earlier, Denver Health also created three high-risk clinics, staffed by multidisciplinary teams specifically designed to meet the needs of patients with the most complex conditions. Some HCIA workforce members worked in different capacities or clinics at Denver Health before the award; others, such as many of the patient navigators, were newly hired at the outset of the intervention period.

Denver Health offered nine training courses for new HCIA-funded staff, including new employee orientation; CMMI orientation (in which participants learned about the purpose, goals, and strategies of 21st Century Care); computer system training; and clinic orientation. Many patient navigators also attended a navigation training session at the University of Colorado, as described in Section 2.d.

Table II.A.2. Key details about intervention staff and workforce development

Program component	Staff members	Staff /team responsibilities	Adaptations?
Enhanced primary care delivery teams/high-risk clinics	Patient navigators	Denver Health placed patient navigators in all eight community health centers. Patient navigators focused on providing between-visit care coordination for patients in risk-stratification Tiers 2–4. Patient navigators also worked in the high-risk clinics recruiting patients and providing care coordination. Patient navigators are not required to have clinical training.	Denver Health refined the role of patient navigators to focus on transitions of care and high-risk care coordination.
	Clinical social workers	Licensed clinical social workers provided systems coordination with outside agencies for the highest-risk children and adults, though the HCIA funded only those social workers serving children. They worked in the community health centers and the high-risk clinics.	None
	Registered nurses	Registered nurses joined all three of Denver Health's general pediatric clinics to provide complex case management for the highest-risk children within the context of the medical home. These nurses also worked in the CSHCN clinic.	None
Enhanced primary care delivery teams	Clinical pharmacists	Denver Health added clinical pharmacists to provide medication therapy management services to high-risk patients, educate providers regarding evidence-based pharmacotherapeutic care for those high-risk patients, and improve medication adherence.	None
	Behavioral health consultants	21st Century Care expanded a previous primary care/behavioral health pilot program that embedded behavioral health consultants in Denver Health's community health centers. Behavioral health consultants worked with patients in need of short-term mental health counseling or other behavioral health needs.	None

Source: Interviews from second site visit, April 2015; document review, March 2015.

Note: This table refers only to staff funded by HCIA. Additional staff positions in the high-risk clinics and community health centers were not funded by HCIA.

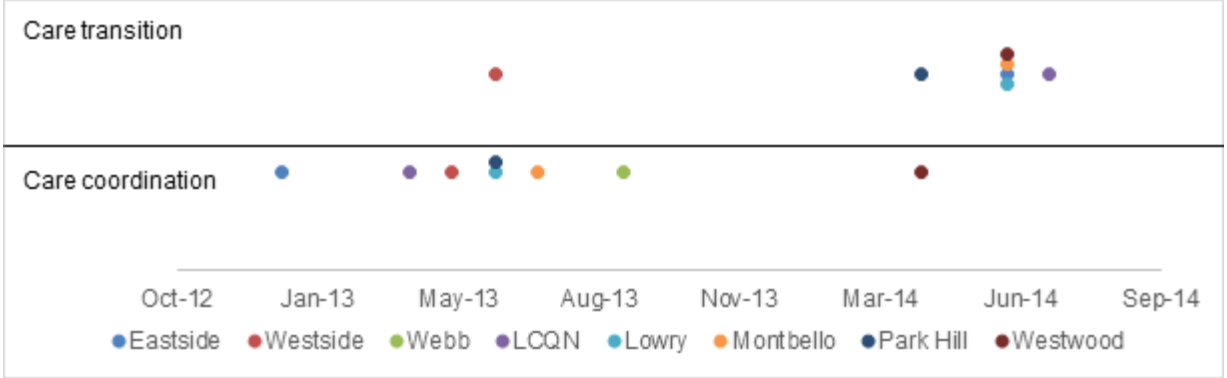
2. Implementation effectiveness

In this section, we examine the evidence of implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators during our second site visits in April, 2015 and self-reported information included in Denver Health's quarterly self-monitoring and measurement reports.

a. Program time line

Denver Health successfully implemented the key aspects of 21st Century Care on schedule. This included implementing the risk-stratification and tiering methodology, redesigning the primary care delivery teams, leveraging health IT to provide between-visit support, and creating three high-risk clinics. However, Denver Health offered its community health centers considerable flexibility in determining when to implement specific components of the 21st Century Care program, particularly the patient navigation activities for care transition and care coordination. Figure II.A.3 shows when each of Denver Health's eight community health centers implemented these two interventions.

Figure II.A.3. Dates of implementation of patient navigation intervention, by clinic



Source: Communication with Denver Health, May 2015.

Note: As of May 2015, Webb Community Health Center had not implemented the transitions of care intervention in its Family Internal Medicine Clinic.

LCQN = La Casa Quigg Newton.

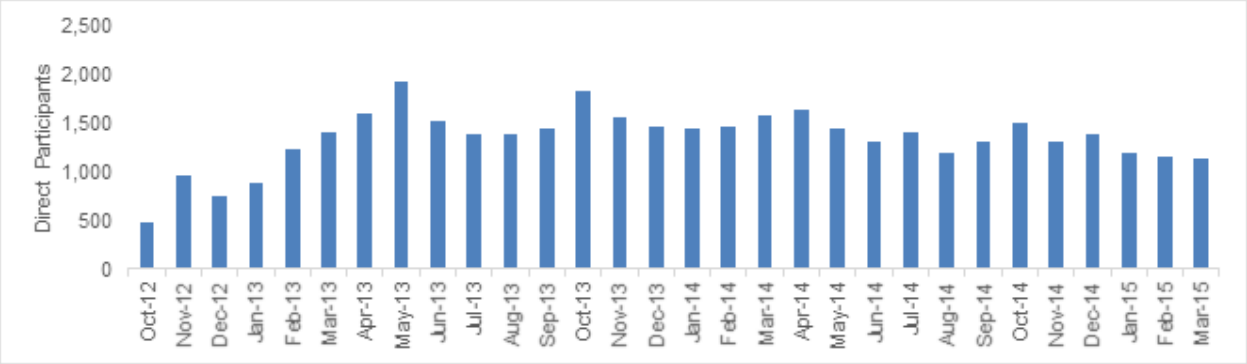
b. Program enrollment

From program inception through December 2014, the 21st Century Care program directly served 16,405 unique patients drawn from the larger Denver Health HCIA program population (see Figure II.A.4 for counts of unique direct participants by month). The number of direct program participants included all patients who received services from a staff member funded under the HCIA, with the exception of patients who received billable services from staff members funded under the HCIA. For example, if an HCIA-funded patient navigator contacts a patient to provide transitions-of-care support, Denver Health counted this as a direct contact,

because the patient navigator was an HCIA-funded position and the service provided was not billable. However, if an HCIA-funded nurse met with a patient during an office visit at the CSHCN clinic, Denver Health counted the office visit as an indirect contact, because this service—despite being provided by an HCIA-funded nurse—was billable. In addition, Denver Health did not include text messaging in its direct patient counts.

Over time, many staff positions funded by the HCIA became operationalized, or transitioned to funding from Denver Health's general operating budget. When this happened, Denver Health reclassified the services delivered by these staff members as indirect. This process occurred in 2014 for behavioral health consultants, and in 2015 for pediatric nurse care coordinators, clinical pharmacists, and IOC and CSHCN staff.

Figure II.A.4. Number of unique direct program participants by month, July 2012–March 2015



Source: Review of Denver Health program reports, March 2015.

Notes: Each bar represents the number of unique participants in that month. Summing two (or more) months would double-count those who participate in two (or more) months.

Six weeks before the end of the program, Denver Health's three high-risk clinics were at or near capacity. The following program benchmarks had occurred as of May 2015:

- The CSHCN clinic had screened and treated 140 high-risk children. Although the CSHCN clinic did not define a specific enrollment goal, staff from the clinic indicated they were near capacity at the time of our site visit.
- The IOC had enrolled and treated 380 high-risk patients, nearly reaching its enrollment goal of 400 patients.
- The MHCD clinic had enrolled 85 patients, close to its capacity of 100 patients.

In addition to direct program participants, program staff invited 103,366 patients to participate in the text messaging intervention as of March 30, 2015. Of these, 23 percent (23,880 patients) enrolled in the service. Participants received on average seven text messages over the course of the intervention. Most text messages were appointment reminders (51 percent), followed by flu vaccine reminders (39 percent), and well-child check reminders (10 percent).

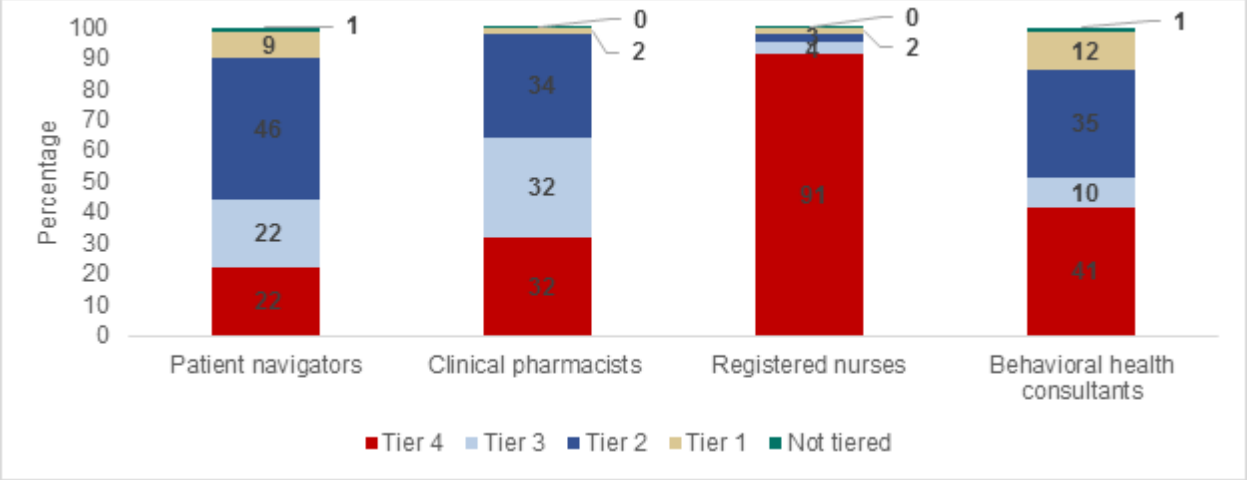
Denver Health also piloted text message programs for diet support and tobacco cessation, but discontinued the tobacco cessation program because patients found it confusing to receive text messages that referred them to a website for tobacco cessation support.

c. Service measures

Figure II.A.5 shows the percentage of direct patient contacts (such as telephone calls, in-person consultations, written letters, and home visits) by type of staff and patients' risk tier. We learned the following about the number and types of patient encounters:

- On average, HCIA-funded staff had one to two contacts with each patient, each month.
- Usually, patient navigators, clinical pharmacists, and behavioral health consultants served patients in Tiers 2, 3, and 4; registered nurses focused on assisting Tier 4 patients.
- Most patient navigators' contact occurred via telephone; 77 percent of all encounters by a patient navigator occurred by telephone compared with only 10 percent through in-person conversations. By contrast, most behavioral health consultants' contact was through in-person conversations (68 percent). Clinical pharmacists and registered nurses relied on telephone conversations, in-person conversations, letters, and other forms of communications.

Figure II.A.5. Percentage of direct patient contacts by staff type and tier, July 2012 – March 2015



Source: Review of Denver Health program reports, March 2015.
Note: Denver Health did not report the number of patient encounters for social workers.

d. Staffing measures

Denver Health was largely successful in hiring intervention staff, exceeding its cumulative new hire full-time equivalent (FTE) target by 55 percent (Lewin 2015). At the height of the program, Denver Health used HCIA funds to support 47.4 FTE staff positions. This included 23 FTE patient navigator positions, 2.8 FTE registered nurse positions, 2.5 FTE clinical pharmacist

positions, and 1.5 FTE clinical social worker positions. Denver Health also used HCIA funding to hire 4.8 FTE clinical and clerical positions for the new high-risk clinics. Finally, Denver Health hired 12.8 FTE administrative, evaluation, and IT staffing positions for the program. The total number of positions supported by the award decreased to 22.6 FTEs in the third year of the award when the start-up health IT roles were eliminated and some staff transitioned from HCIA to general operational funding.

Denver Health experienced some staffing challenges, particularly with patient navigators and IT staff. Identifying and retaining people who were a good fit for the position was a challenge. Denver Health administrators, clinicians, and staff reported a high degree of turnover among patient navigators and added that many high-performing navigators used the position as a stepping stone to more formal training in the medical field. Over the course of the award period, Denver Health hired 51 patient navigators for its 24 positions. Denver Health administrators also reported problems finding qualified IT staff and vacant positions in the IT department were a persistent problem throughout the award.

Denver Health achieved its total trainee projection targets. Intervention staff completed more than 9,311 hours of training from July 2012 to March 2015 (Lewin 2015). Twenty-one HCIA-funded staff also participated in a 32-hour patient navigation training certification course. Denver Health developed the course in collaboration with the University of Colorado Health Sciences Center; the School of Public Health at the University of Colorado administered the course. The curriculum covered knowledge and skills related to patient navigators' core competencies.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) implementation process, (2) program characteristics, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.3 summarizes the major facilitators and barriers to Denver Health's implementation effectiveness in each domain.

Table II.A.3. Facilitators and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Empowered frontline staff can adapt implementation strategies and activities based on patients' needs • Commitment to staff collaboration across multidisciplinary teams	• Challenges posed by small multidisciplinary teams, including personality clashes and unfilled positions
Implementation process	• System wide emphasis on self-monitoring and continuous quality improvement	• Integration of patient navigators into care teams • Engagement of Denver Health providers not directly involved in 21st Century Care
Internal factors	• Buy-in by providers and staff to the integrated care team model • Supervision and support of new clinical staff	• Lack of an interoperable, system-wide electronic health record (EHR)
External environment	• Payment model reform in Colorado allowed for billing of integrated behavioral health consultant visits, which allowed these roles to become self-sustaining	• Complexity of patients' needs and patients' resource constraints

Source: Interviews from second site visit, April 2015; document review, March 2015.

a. Program characteristics

Two key characteristics of 21st Century Care had a major impact on implementation of the program. First, Denver Health empowered its frontline staff to adapt implementation strategies and activities based on the needs of their patients. For example, frontline staff in the IOC reported they expanded services to better meet patients' needs, including the addition of group visits for pain management and hospital rounding by physicians. In addition, frontline staff continually refined their patient identification processes. For instance, staff at the CSHCN clinic reported they developed new patient lists to proactively identity CSHCN with gaps in care. Another example of 21st Century Care's empowerment of frontline staff was the CSHCN staff's decision to travel each month to the pediatric clinics to support primary care providers with the treatment and care of CSHCN. This decision enabled patients to remain in their medical homes while also receiving more intensive and targeted care from the CSHCN staff.

Second, an important aspect of the new high-risk clinics was staff members' commitment to collaborate across multidisciplinary teams. This approach to care enabled patients to see multiple professionals during one visit. Because patients at high-risk clinics often have serious barriers to accessing care—including transportation and mental health issues—staff reported that one-stop shopping for medical, behavioral, and social services improved patients' overall care compliance. However, providers and staff noted that due to the small size of the teams, personality clashes or unfilled positions resulted in significant challenges in the high-risk clinics.

b. Implementation process

Denver Health's system wide emphasis on using Lean methods of self-monitoring and continuous quality improvement facilitated its implementation of 21st Century Care. Two factors—the integration of new staff into care teams and the engagement of Denver Health providers not directly involved in 21st Century Care—initially posed challenges to implementation, but over the course of the award period staff overcame these challenges.

Throughout the implementation of 21st Century Care, Denver Health used Lean methodology to improve care. The Lean methodology offers a process and management improvement system that relies on self-monitoring, continuous quality improvement, and the elimination of waste. Using the Lean methodology involved holding frequent rapid-improvement events with 21st Century Care team leaders and frontline staff to refine staff roles, improve processes, and redesign workflows. Staff reported the use of Lean processes encouraged input from people in different roles throughout the Denver Health system and facilitated system wide improvements in a more immediate way than otherwise might have been possible. For example, Denver Health held iterative Lean events over the course of the HCIA award that focused on developing the enhanced care teams. These Lean events enabled Denver Health to quickly adapt care team redesign plans when faced with changes in management structure, continue learning about how best to use various staff roles, and increase knowledge of the needs of the high-risk patient population. In addition, Denver Health used Lean processes to pilot test activities in a few community health centers to identify strategies to possibly implement on a larger scale. For example, Denver Health pilot tested the high-risk/high-cost care coordination intervention to assess different strategies related to patient selection, frequency of case conference meetings, assignment of responsibilities across medical positions, and care plan documentation.

Initially, Denver Health faced challenges integrating new staff, especially patient navigators, into care teams because existing staff had limited or no experience working with patient navigation. Moreover, the patient navigators' role within the care teams was not clearly defined at the beginning of the program. For example, patient navigators initially led case conferences in some clinics, even though care team members reported that patient navigators did not have the appropriate training and skills to lead these comprehensive evaluations of patients' needs. Clinic staff reported that Denver Health improved the definition of patient navigators' roles and responsibilities, leading to greater integration of these new staff into the care teams. Patient navigators shifted from leading case conferences to assisting nurses and social workers with patients' assessments. Denver Health discovered that patient navigators were also well equipped to work with patients on care transitions and to help them with transportation and appointment scheduling needs. Staff reported that greater clarity about the role of the patient navigator led to more focused and effective work by the patient navigators, which helped providers and other staff reduce the amount of time they had to spend on nonclinical tasks.

Another area of improvement was an increased awareness of the new high-risk clinics among providers and staff throughout the Denver Health system, including those not funded by the HCIA. Periodic outreach by IOC and CSHCN clinic staff—including activities such as high-risk clinic team leaders speaking at Denver Health staff meetings and distribution of written and electronic materials describing the clinics and referral processes—aided in the process of

introducing the new clinics to the system as a whole. As more providers and staff referred patients to the high-risk clinics, they also became more familiar with the array of special services the clinics could provide to patients with the most complex needs. For example, Denver Health providers and staff reported that the IOC and the CSHCN clinics were valuable resources for high-risk patients who needed more time with a care team than the typical 15- or 20-minute primary care visit. In addition, Denver Health providers and staff appreciated that IOC and CSHCN clinic patients received multiple medical, behavioral, and social service needs in one visit. Providers and staff also reported the specialty high-risk clinics helped reduce stress in the regular primary care settings by enabling providers to refer challenging patients who would be better served in the high-risk clinics.

c. Internal factors

Three internal factors affected implementation effectiveness of the 21st Century Care program. First, widespread provider and staff buy-in into the integrated care team model facilitated implementation of 21st Century Care. Primary care providers and staff expressed support for continued involvement of patient navigators, behavioral health consultants, social workers, and clinical pharmacists as critical members of the care team. Care teams working in the IOC and CSHCN clinics reported being particularly interdependent, and emphasized that teamwork, flexibility, communication, and willingness to step outside of their usual roles to help patients were key factors in the successful implementation of team-based care in these clinics.

Second, strong supervisory and peer support for new staff facilitated program implementation. Behavioral health consultants and social workers reported that meetings with their supervisors and peers outside the clinic were particularly helpful for building moral support and solving problems. Several providers and staff members interviewed mentioned that stronger central oversight and supervision of patient navigators would have improved the implementation processes of 21st Century Care; however, most providers and staff added that the supervision of patient navigators had improved significantly over the course of the award period.

Denver Health's lack of an interoperable, system-wide EHR was a barrier to implementation of 21st Century Care. Clinicians and staff reported that the use of multiple systems for tracking 21st Century Care activities created data and communication challenges for the integrated care teams. Denver Health expects to roll out a new EHR system in 2016.

d. External factors

Features of an organization's external environment can also influence program implementation. Two external factors affected program implementation. First, the introduction of Medicaid payment model reform in Colorado allowed Denver Health to bill for same-day, integrated behavioral health visits. As a result, starting in July 2014, this new source of revenue enabled the behavioral health consultants to become self-sustaining members of the care teams.

The characteristics of 21st Century Care's target population posed implementation challenges. Denver Health serves primarily low-income patients who face multiple barriers to health and health care, such as language barriers, financial impediments, transportation issues, and low health literacy and knowledge. Although Denver Health designed 21st Century Care to

reach this population and was aware of potential challenges, certain patient barriers emerged as being particularly problematic. For example, patient navigators reported that it was difficult to contact patients because many had pay-by-the-minute cell phones and the numbers constantly changed, or because patients who were homeless had no means of communicating, other than showing up in person at a clinic.

4. Sustainability and scalability

By May 2015, Denver Health had successfully incorporated two-thirds of all HCIA-funded staff into its internal operational budget. The transfer of behavioral health consultants occurred in July 2014. In January 2015, Denver Health started paying for pediatric nurse care coordinators, clinical pharmacists, and IOC/CSHCN staff through its internal operational budget. Denver Health also eliminated several positions, such as the health IT start-up positions, which were always intended to be short-term positions. As of May 2015, Denver Health administrators reported facing challenges to transitioning patient navigator positions to internal funding; these positions account for the remaining one-third of program personnel costs, but Denver Health has not identified a way to fund them. Internal evaluation efforts were underway to estimate the impact of patient navigator-led interventions on targeted populations.

Denver Health has already begun spreading 21st Century Care program activities to additional clinics. For example, a new patient navigator now conducts transitional care outreach calls for the Denver Health geriatric clinic, and the system's HIV clinic is beginning to adapt this intervention for its patients using a patient navigator not funded by the HCIA. The HIV clinic is also implementing care coordination, modeled after the intervention developed for 21st Century Care.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected clinical sites or satellite offices provided important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff are limited to a small number of clinical locations and might not reflect the perspectives of clinicians practicing at other sites. In order to assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in the fall of 2014, the third year of the HCIA-funded program.The Denver sampling frame included all Denver Health clinicians involved in the 21st Century Care program. There were 140 clinicians in the sample file sent to us by Denver Health. Data from the survey provide additional insights into the implementation process and experience as well as the contextual factors that might affect implementation effectiveness at Denver Health.

In this section, we report on Denver Health clinicians' views of their daily work life and practice. First, we focus on the contextual factors that can affect program implementation, including the characteristics of the practice locations, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care teams function. We then present data on the alignment of Denver Health

clinicians' views and experiences with the overall goals of the HCIA-funded innovation, as well as their awareness of and participation in the PCMH program and their view of the facilitators of and barriers to successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

A total of 81 Denver Health clinicians responded to the survey (resulting in a response rate of 67 percent). Of the respondents, 59 were physicians, 10 were nurse practitioners, and 12 were physician assistants. Denver Health clinicians practice at FQHCs. Most Denver Health clinicians reported that their primary source of compensation was a fixed salary (81 percent).

Denver Health clinicians reported mixed uses of health IT, which reflects that Denver Health does not currently have a system wide EHR, but will introduce a new EHR in April 2016. At least 90 percent of Denver Health clinicians used electronic systems to access laboratory test results, to enter clinical notes, for referral tracking systems, and for patient registries. However, fewer than 20 percent of clinicians reported using electronic systems to order tests and procedures, for prescribing, or for alert warnings of drug dosing or drug interactions.

Denver Health clinicians also did not report high levels of patient-facing technologies, such as allowing patients to request an appointment online (15 percent), allowing patients to email a clinician about a medical question or concern (28 percent), or allowing patients to request a prescription refill online (38 percent). As part of 21st Century Care, Denver Health determined that offering patient-facing technologies that relied on Internet access did not best meet the needs of their patient population, choosing instead to use text messaging to reach them, because a higher percentage of patients had access to this technology.

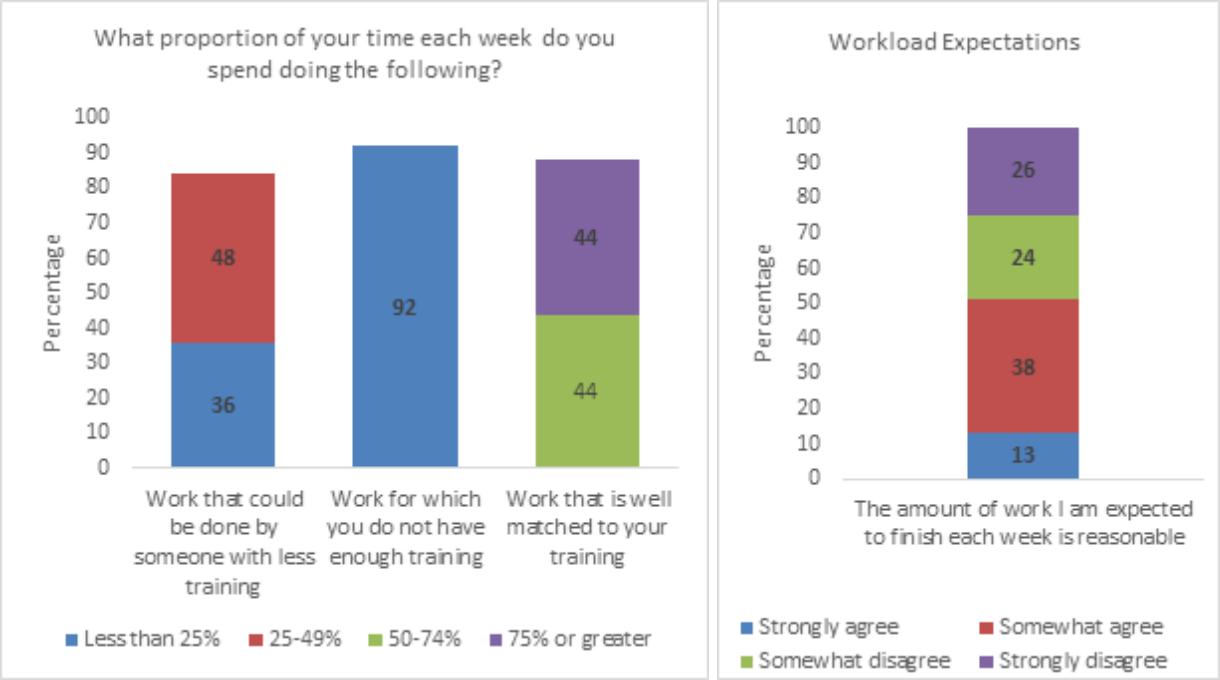
b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall career, level of burnout, and perceptions of their practice environment can all have an effect on the success of program implementation and organizational change. Denver Health clinicians are generally satisfied with their careers in medicine (91 percent were at least somewhat satisfied). Most clinicians agreed at least somewhat that their management team was supportive (89 percent), that they were encouraged to offer suggestions and improvements (90 percent), and that they had adequate opportunities for professional development (84 percent). Overall, most Denver Health clinicians reported only occasionally feeling burned out (52 percent), though 38 percent were experiencing one or more symptoms of burnout at the time the survey was taken.

One key aspect of 21st Century Care was the addition of new staff members, such as patient navigators and behavioral health consultants, to Denver Health's community health centers. These new staff were intended to help clinics expand service offerings and to shift work so that each member of the care team could provide services appropriate to his or her training. Most Denver Health clinicians reported that their work matched well to their training and agreed that they had enough training to support their work (Figure II.B.1). However, almost half of clinicians thought that 25 to 49 percent of their work could be done by someone with less

training. In particular, physicians and nurse practitioners felt that they performed a significant amount of work that someone with less training could do. Clinicians had mixed feelings on whether the amount of work they were expected to complete each day was reasonable, with 26 percent of clinicians strongly disagreeing that their workload was reasonable.

Figure II.B.1. Workload, training, and expectations



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

Denver Health clinicians also varied in their opinions of the quality of care they are able to provide, with only half agreeing at least somewhat that they were able to provide high quality care to all patients. As shown in Table II.B.1, all respondents who believed that the question applied to their practice felt that a major barrier to providing optimal care was having enough time to spend with patients during visits. Other major barriers to care that a majority of clinicians reported were insufficient reimbursement, patients' inability to pay for care, lack of timely information about care provided to patients by other physicians, difficulties obtaining specialized diagnostic tests or treatments, and difficulties obtaining specialist referrals. Most Denver Health clinicians did not report lacking adequate information from research evidence to guide their clinical decisions.

Table II.B.1. Perceptions of ability to provide high quality care to all patients

Survey item	Number of respondents	Percentage of respondents
Percentage reporting each of the following at least somewhat limits their ability to provide optimal, patient-centered care		
I do not have enough time to spend with patients during visits.	80	100%
The level of reimbursement is not adequate.	68	87%
My patients have difficulty paying for needed care.	65	82%
I lack timely information about the patients I see who have been care for by other physicians.	63	82%
It is difficult for me to obtain specialized diagnostic tests or treatments for my patients in a timely manner.	54	69%
It is difficult for me to obtain specialist referrals for my patients in a timely manner.	51	64%
I lack adequate information from research evidence to guide my clinical decisions.	23	29%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

c. Clinicians' perceptions of care team functioning

Ninety-three percent of Denver Health clinicians reported working as part of a care team. Overall, their perceptions of how these teams function were positive. Most Denver Health clinicians agreed that members of the care team relayed information in a timely manner (90 percent), used common terminology when communicating with one another (88 percent), had sufficient time for patients to ask questions (81 percent), verbally verified information they received from one another (76 percent), and followed a standardized method of sharing information when handing off patients (63 percent).

d. Clinician engagement in other quality improvement activities

Denver Health clinicians reported being highly involved with quality improvement activities over the past two years. As described earlier, Denver Health uses Lean methodology, a process and management improvement system that relies on self-monitoring, continuous quality improvement, and the elimination of waste. Eighty-two percent of clinicians reported receiving training on quality improvement tools and conducted at least one clinical audit of patients' care. In addition, about half of clinicians participated in collaborative quality improvement activities with other practices, hospitals, government agencies, or professional associations (52 percent).

e. Alignment with goals of HCIA program

Clinicians were asked to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. In Table II.B.2, we present results based on the proportion of clinicians rating each of these goals as extremely important. The inclusion of the extremely important category helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are simply important.

Table II.B.2. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Improving patients' capacity to manage their own care	52	64%
Increasing access to primary care	50	62%
Improving care coordination for patients with chronic conditions	48	59%
Reducing hospital readmissions	45	56%
Reducing ED visits	43	53%
Improving appropriateness of care	42	52%
Reducing overall health care spending	40	49%
Improving care continuity in primary care	38	47%
Increasing use of EHRs and other health IT	38	47%
Improving capability of health care organizations to provide team-based care	32	40%
Improving the capability of health care organizations to provide patient-centered care	31	38%
Increasing the use of evidence-based practice in clinical care	29	36%
Increasing the number of primary care practices functioning as PCMHs	28	35%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Number of respondents are those who rated the following as extremely important.

The views of Denver Health clinicians generally aligned with the goals of 21st Century Care. Most clinicians in the sample reported that improving patients' capacity to manage their own care, increasing access to primary care, improving care coordination for patients with chronic conditions, reducing hospital readmissions and ED visits, and improving appropriateness of care were extremely important goals.

3. Awareness of program and perceived effects

Understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on patients' outcomes. For example, if clinicians are aware of the program, have received appropriate and effective training, and believe that 21st Century Care will have a positive effect on the care they provide, they are likely to feel more invested in the program's success. Alternatively, those who feel more negatively about the program might be less likely to implement the intervention enthusiastically. In this section, we report on Denver Health clinicians' experiences with and perceptions of the 21st Century Care program.

a. Awareness of program and perceived effects

Eighty-four percent of the Denver Health clinicians we surveyed were at least somewhat familiar with the 21st Century Care program. Overall, perceptions of 21st Century Care were positive. Most clinicians who were familiar with the program believed it would have a positive effect on the quality of care (68 percent), improve patient-centeredness (69 percent), and

improve their ability to respond in a timely way to patients' needs (66 percent). Fewer than half of physicians familiar with the program believed it would have a positive effect on equity, efficiency, and safety. Very few clinicians perceived a negative impact of the program; rather, some believed the intervention would have no effect on the care they provide or that it was simply too soon to tell.

b. Barriers to and facilitators of program implementation

Finally, Denver Health clinicians who were at least somewhat familiar with 21st Century Care were asked to rate the effect of a series of barriers to and facilitators of program implementation. The top three reported facilitators to program implementation were the availability of personnel (66 percent), level of funding (56 percent), and availability of community resources to care for patients with complex conditions (54 percent). The most often-cited barrier to program implementation was the amount of required documentation, which 19 percent of responding clinicians cited.

4. Conclusions about clinicians' attitudes and behavior

Most clinicians surveyed were aware of 21st Century Care and believed it would have a positive effect on quality of care, patient centeredness, and their ability to respond in a timely way to patient needs. In particular, clinicians believed that the availability of personnel, level of funding, and availability of community resources to care for complex patients were key factors in implementing the program. Survey results suggest that contextual factors may have had a mixed impact on the implementation of 21st Century Care. For example, clinicians in the Denver Health sample reported moderate levels of clinician burnout, and generally positive attitudes toward practice management and working in care teams. These positive contextual factors may have helped facilitate the implementation of the program. On the other hand, only half of Denver Health clinicians believed they could provide high quality care to all their patients. Clinicians reported the major barrier to providing optimal care was having enough time to spend with patients during visits, a problem that could be related to the fact that most clinicians reported spending a significant amount of time on work that could be done by someone with less training.

C. Impacts on patients' outcomes

1. Introduction

In this part of the report, we present preliminary results about the impacts of Denver Health's HCIA program on patients' outcomes in two domains: service use and spending. Results are preliminary because our analyses do not yet cover the full time period that we will include in the final impact analysis. Moreover, we have not yet drawn conclusions about program impacts during the early period of the award because our initial results, described later, suggest we might need to refine the sample or statistical model used for the impact analysis before we can draw conclusions. The preliminary results in this report should therefore be considered a work in progress. Finally, although the Denver Health program serves a largely Medicaid or uninsured population, due to limitations in available data we have analyzed outcomes only for the Medicare fee-for-service (FFS) population (including those who are dually

eligible for Medicare and Medicaid). Results might not be generalizable to the full population that Denver Health's program serves.

This part of the report is organized into four sections, in addition to this introduction. First, we describe the methods we have used so far for estimating impacts (Section II.C.2) and then the characteristics of the Denver Health Medicare FFS beneficiaries who form our treatment group for the impact analysis (Section II.C.3). We next demonstrate that, before they received any HCIA-funded services, the 6,199 treatment group members were similar to the 23,196 people we selected as a comparison group (Section II.C.4). Finally, in Section II.C.5, we present preliminary quantitative estimates, discuss their agreement with implementation findings, and describe our next steps for the impact evaluation.

In our future analysis, we will conduct additional sensitivity tests, assess impacts over the full period of the Denver Health HCIA, and draw conclusions about program impacts for Medicare FFS beneficiaries. We might also include Medicaid FFS beneficiaries, pending available Medicaid claims data.

2. Methods

a. Overview

We have generated preliminary impact estimates as the difference in outcomes for people assigned to the treatment group and outcomes for people matched as a comparison while (1) adjusting for any individual-level characteristics that might affect outcomes and (2) subtracting out any differences in outcomes that existed between the groups during an 18-month baseline period before 21st Century Care began. This analytic strategy is called a difference-in-differences model because it estimates impacts as the change in the difference between the treatment and comparison groups that occurred between the baseline period (before the HCIA award began) and the intervention period (after the HCIA award began).

To focus our impact analyses, we specified a limited number of primary tests before conducting the analyses. Each primary test defined an outcome, population, time period, and direction of expected effects for which we hypothesize impacts if the program is effective. We provided both Denver Health and CMMI an opportunity to comment on the primary tests. We developed a set of decision rules to draw conclusions about impacts in each domain (service use or spending) based on the results of these primary tests and the consistency of the primary test results with the implementation findings and with secondary quantitative tests (robustness and model checks).

b. Treatment group definition

We defined the Medicare FFS treatment group separately in each of six baseline quarters before 21st Century Care began (the baseline period), and in each of nine intervention quarters after the program began (the intervention period). We defined the first intervention quarter (I1) to begin on November 1, 2012, which was the first day of the first month after the 21st Century Care start date (October 29, 2012). Later in this report we present detailed information on the treatment group construction. However, in brief, we defined the treatment group in each quarter

to consist of Medicare FFS beneficiaries (1) identified by the awardee as meeting the program eligibility criteria outlined in Section I.A.1.b *during the period* (baseline or intervention) on or before the first day of the quarter; (2) with 12 months of continuous enrollment in Medicare FFS before they began receiving services under the 21st Century Care program (to facilitate matching to potential comparison beneficiaries); and (3) with observable outcomes for at least one day in the quarter. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer. This includes beneficiaries who are dually eligible for Medicaid.

Using this definition, a beneficiary who has previously been assigned to the treatment group will remain a member of the treatment group for the rest of the relevant period (baseline or intervention), as long as he or she is still enrolled in Medicare FFS. This definition ensures that, during the intervention period, beneficiaries do not exit the treatment group solely because the intervention succeeded in reducing their service use, including visits at Denver Health. The definition for the baseline period corresponds to that of the intervention period so that, across the two periods, there should be broadly similar changes in population composition.

Identification of Medicare FFS beneficiaries from Denver Health data. Denver Health provided records for 251,477 distinct program participants who met the 21st Century Care program eligibility criteria at some point from May 1, 2011, to January 31, 2015. As described in Section I.A.1.b, these people either (1) had made a primary care visit to a Denver Health clinic in the previous 18 months; (2) had made three or more visits to Denver Health's acute care facilities (the hospital, the ED, or an urgent care clinic) in the previous 12 months or had had two hospitalizations during this period and a qualifying mental health diagnosis; or (3) were members of a Denver Health managed care plan. This number of program participants—251,477—is larger than the one shown in Figure II.A.2 because that figure shows program participation at a single point in time during the intervention period, whereas the records used for the impact analysis cover all patients meeting the criteria at any time during either the baseline or intervention periods. Due to a lag in our data processing for the comparison group (described in Section II.C.2.c), we excluded 19,832 participants who visited Denver Health for the first time after June 2014 because we could not match them to comparison beneficiaries. We then merged the Denver Health data with Centers for Medicare & Medicaid Services (CMS) data on Medicare enrollment, identifying 19,514 Denver Health program participants who were enrolled in Medicare (either FFS or Medicare Advantage) at some point during the baseline or intervention periods.

Assignment to the treatment group in each month and each quarter. We assigned beneficiaries on a monthly basis. We then used the monthly assignments to assign each beneficiary in each quarter because we analyzed impacts on quarterly outcomes in our preliminary impact analysis.

For the monthly assignments, we assigned each Medicare beneficiary to the treatment group starting in the first month in which he or she both met the Denver Health program eligibility criteria and had 12 months of continuous observability in Medicare FFS claims before receiving 21st Century Care services. This meant we restricted the treatment group to Medicare FFS

beneficiaries who either (1) met the Denver Health eligibility criteria at some point during the baseline period and were observable for 12 months in FFS claims either immediately before the first month in which they met the criteria, or in a subsequent month before the intervention began (because it was not possible to receive intervention services before the intervention start date); or (2) in the case of beneficiaries who made their first visit to Denver Health after the program start date, met the Denver Health eligibility criteria at some point during the intervention period, and were observable in Medicare FFS claims data for 12 months before their first visit at Denver Health. This restriction resulted in a final treatment group of 6,199 Medicare FFS beneficiaries.

We then used the monthly treatment group assignments to determine whether each beneficiary was assigned in each quarter (because we analyzed impacts on quarterly outcomes). Specifically, within each period (either the baseline or intervention period) we considered a beneficiary assigned to the treatment group in a given quarter if he or she was assigned to the treatment group in or before the first month of the quarter.

As noted earlier, when beneficiaries are assigned to the treatment group during a period, they remain assigned for the duration of the period. For example, assuming 12 months of prior Medicare FFS enrollment, a beneficiary who first met the Denver Health eligibility criteria in November 2012 (the first month of the intervention period) would be assigned to the treatment group for the entire intervention period. A beneficiary who met the Denver Health eligibility criteria only during the baseline period would be assigned to the treatment group for the duration of the baseline period following initial quarterly assignment, but would not be assigned during the intervention period. A beneficiary meeting the criteria in both periods would be assigned in both.

Definition of high-risk subgroup. Because the 21st Century Care program stratified patients into risk tiers and offered additional services to high-risk patients, we further defined a high-risk subgroup of the overall treatment group. We used Medicare claims to identify beneficiaries who, when they first joined the treatment group, met Denver Health's criteria for frequent users of acute care facilities (which, as described in Section I.A.1.b, is one way a beneficiary can meet program eligibility criteria). As noted earlier, a frequent user is a beneficiary with three or more acute care visits (hospitalizations, ED visits, or urgent care visits) in a year, or two or more hospitalizations in a year and a qualifying mental health condition. This definition of *high risk* does not map directly to the awardee's risk tiers and is more inclusive than Denver Health's highest-risk population, Tier 4. By using the more inclusive definition, we aim to assess the impact of Denver Health's tiering algorithm (which evolved over time) as a tool to allocate resources among a larger and more broadly defined at-risk population. We also aim to capture the impacts of HCIA-funded services for some Tier 2 and 3 patients—for example, reorganization of primary care delivery teams and the newly hired patient navigators.

When a person is assigned to the high-risk subgroup, he or she remains a member of that group as long as assigned to the treatment group. This prevents people from exiting the high-risk subgroup used in our analysis solely because the intervention succeeded in lowering their risk.

c. Comparison group definition

In each baseline and intervention quarter, the comparison group consists of Medicare FFS beneficiaries who were matched to a treatment beneficiary based on demographic and health characteristics. Because there are no health systems similar enough to Denver Health (in both population served and regulatory environment) to serve as an appropriate comparison group, we constructed our comparison group by matching individual treatment beneficiaries to individual comparison beneficiaries. We selected these comparison beneficiaries in three steps.

First, we selected a pool of potential comparison beneficiaries, which consisted of all Medicare FFS beneficiaries who were observable and living in the state of Colorado, but outside the City and County of Denver, for at least one month from May 2011 to June 2014. (May 2011 was the start of the baseline period.) We excluded beneficiaries living in Denver from the pool because those who live in the city but do not receive care from Denver Health could differ systematically and in unobservable ways from those in the treatment group. We used Medicare claims and enrollment data to construct a person-month file for the potential comparison beneficiaries—meaning that the file contained one record for each month that a beneficiary was observable and living in the designated geographic area. We then restricted this beneficiary-month pool to include only those monthly records in which the beneficiary met either of two Denver Health program eligibility criteria related to service use: (1) having had a primary care visit in the previous 18 months; or (2) having had three or more acute care visits (that is, hospitalizations, ED visits, or urgent care visits) in the past 12 months or two or more hospitalizations and a qualifying mental health diagnosis. This definition allows the comparison pool to include beneficiaries who could be good matches for treatment beneficiaries who entered the treatment group through either the primary care or acute-care eligibility pathway. (Beneficiaries who were part of a managed care plan—the third eligibility pathway—are typically not observable in Medicare FFS claims.) Finally, for a beneficiary-month observation to be included in the comparison pool, we required the beneficiary to have been continuously enrolled in FFS Medicare for the prior 12 months. (This restriction facilitates matching by ensuring that our matching variables, developed by analyzing claims, are reliable.) This left a pool of 322,104 distinct Medicare FFS beneficiaries, most of whom contributed multiple observations (that is, multiple months) in which they could potentially be matched.

Second, we developed matching variables for all treatment group members (defined as of the month each beneficiary was first assigned to the treatment group) and all potential comparison beneficiary person-months (that is, all person-months meeting the program eligibility criteria and having at least 12 months of previous continuous Medicare FFS enrollment). These matching variables included indicators of health care utilization and risk, including past ED visits and hospitalizations; past Medicare spending; Medicare/Medicaid dual eligibility; Hierarchical Condition Category (HCC) scores (which reflect estimated spending in the coming year); whether the beneficiary qualified for the high-risk subgroup in that month; and original reason for Medicare eligibility (either old age, disability, end-stage renal disease [ESRD], or both disability and ESRD). The matching variables further included demographic characteristics, including age, sex, race or ethnicity, and socioeconomic characteristics of the zip code in which the beneficiary resided. Section II.C.4 describes the matching variables in detail.

Third, we used the matching variables to match the 6,199 treatment beneficiaries to 23,196 comparison beneficiaries (from the pool of 322,104). This section describes how we matched the matched comparison group, whereas Section II.C.4 presents the balance we achieved between the two groups on the matching variables. We started by matching the treatment group members assigned to the treatment group in the first month of the first baseline quarter (May 2011). We then matched each monthly cohort of newly assigned beneficiaries separately. This means, for example, that we matched every treatment group member assigned in May 2011 to one or more comparison beneficiaries in that same month (that is, using the May 2011 observations in the comparison pool file), before proceeding to match the treatment group members assigned for the first time in June 2011. Each treatment beneficiary was matched to up to five beneficiaries from the potential comparison pool to increase the statistical certainty in the impact estimates (relative to 1:1 matching). We allowed comparison group members to match no more than once, meaning that comparison beneficiaries who matched in May 2011, for example, were removed from the pool of potential comparison group members for all subsequent months.

To execute this matching, we used a combination of exact-matching, caliper-matching, and propensity scores. *Exact-matching* means that we forced a treatment beneficiary to have an identical value of a given variable to his or her matched comparisons. We exact-matched on whether the beneficiary was a member of the high-risk subgroup. *Caliper-matching* means we forced each treatment beneficiary to have a value for a given variable that fell within a specified range. We caliper-matched on age, requiring that each treatment beneficiary match to comparison beneficiaries no more than five years older or younger. For all other variables we matched using *propensity scores*. A propensity score is the predicted probability, based on all of a beneficiary's matching variables, that a given beneficiary was selected for treatment (Stuart 2010). In other words, it collapses all of the matching variables into a single number for each beneficiary that can be used to assess how similar beneficiaries are to one another. By matching each treatment beneficiary to one or more comparison beneficiaries with similar propensity scores, we generated a comparison group that is similar, on average, to the treatment group. We used separate propensity-score models for each monthly cohort of beneficiaries, and, within each month, we used separate propensity-score models for the high-risk beneficiaries and other (not high-risk) beneficiaries.

d. Restriction of sample to complete matched sets

After matching, we restricted the sample for analysis in each quarter to complete matched sets—that is, to groupings of treatment and matched comparison beneficiaries in which all members were observed in Medicare FFS claims for the quarter. (For example, if a matched set were assigned in the first baseline quarter (B1) but one of the comparison beneficiaries died by the second baseline quarter, we would include that matched set in the analytic sample for the first quarter but not for subsequent quarters.) This sample restriction limits potential bias due to differential attrition between the treatment and comparison groups. Descriptive statistics (not shown) suggested that treatment beneficiaries were more likely than comparison beneficiaries to switch from Medicare FFS into a Medicare Advantage managed care plan and, furthermore, that the beneficiaries who switched into managed care had below average service use and spending before switching. Thus, the sample restriction limits the likelihood that relatively healthy

treatment beneficiaries exit the sample (making the treatment group less healthy overall) while their matched comparison beneficiaries remain in the analytic sample.

e. Construction of outcomes and covariates

We used Medicare claims from May 1, 2008, to January 31, 2015, for beneficiaries assigned to the treatment and comparison groups to develop two types of variables: (1) **outcomes,** defined for each person in each baseline or intervention quarter during which they were a member of the treatment or comparison group; and (2) **covariates** that describe a beneficiary's characteristics in the month in which they were matched, and that we used in the regression models for estimating impacts to adjust for beneficiaries' pre-intervention characteristics. We used covariates defined before each beneficiary was first assigned to the treatment or comparison group, without updating them each quarter, to avoid controlling in each intervention quarter for previous quarters' program effects, as this would bias the effect estimates away from detecting true impacts. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. We calculated three quarter-specific outcomes that we grouped into two domains:

1. Domain: Service use
    a. All-cause inpatient admissions (number/quarter)
    b. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
2. Domain: Spending
    c. Total Medicare Part A and B spending ($/month)

These outcomes—along with the 30-day unplanned hospital readmission rate—are outcomes that CMMI has specified as core for the evaluations of all HCIA programs. We did not include 30-day unplanned readmissions among our primary tests because Denver Health did not state explicitly in its HCIA application that it intended to affect this outcome. The primary tests assess impacts only on those outcomes for which we most strongly expect to detect effects if the program was indeed effective.

Covariates. The covariates include (1) demographic factors (age, gender, race/ethnicity) and socioeconomic characteristics of the zip code in which the beneficiary resided (unemployment rate, proportion of people living below the federal poverty level, and proportion with a college degree); (2) original reason for Medicare entitlement (old age, disability, ESRD, or both disability and ESRD); (3) Medicare and Medicaid dual-eligibility status; (4) HCC scores; (5) whether the beneficiary had each of 22 chronic conditions (including physical health, mental health, and disabilities), created by applying Chronic Conditions Warehouse algorithms to claims data in the 12 to 36 months (depending on the condition) before entry into the analytic sample; (6) measures of recent service, including the numbers of hospital admissions, outpatient ED visits, and 30-day unplanned readmissions in the 12 months before entry into the analytic sample, and the number of primary care visits in the preceding 18 months; (7) total Medicare

Part A and B spending in the 12 months before entry into the analytic sample; (8) whether the beneficiary qualified as having a serious mental health diagnosis, as defined under Denver Health's 21st Century Care program eligibility criteria; and (9) for the analyses among the full Medicare FFS population (but not among the high-risk subgroup) whether the beneficiary was a frequent user of acute-care services, following the definition from Denver Health's program eligibility criteria. In addition, the covariates included one variable that was not defined as of the first month the beneficiary was assigned, but instead captured whether the beneficiary *ever* visited an FQHC participating in CMMI's FQHC Demonstration from November 2011 (when the demonstration started) to June 2014. This variable is defined differently from other covariates for reasons described in Section II.C.2.h.

f. Regression model

We used a regression model to implement a difference-in-differences design for estimating impacts. For each outcome in each quarter, the model estimates the relationship between the outcome and a series of predictor variables, assuming that each one of the predictor variables has a linear (additive) relationship with the outcome. The predictor variables include the beneficiary-level covariates (defined in Section II.C.2.e), an indicator (fixed effect) for each matched set in each quarter to absorb the mean outcome for the matched set in the quarter, an indicator for whether the beneficiary is assigned to the treatment or comparison group, and an interaction of a beneficiary's treatment status with each post-intervention quarter. The estimated relationship between the interaction term and the outcome in a given quarter is the impact estimate for that outcome in that quarter. It measures the average difference between outcomes for beneficiaries assigned to the treatment and comparison groups that quarter, subtracting out the average difference between these groups during the six baseline quarters. By providing separate impact estimates for each intervention quarter, the model enables the program's impacts to change the longer the program is underway (which is what the awardee expected). We can also test impacts over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the impact estimates, allowing for statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. Appendix 2 provides details on the regression methods, including descriptions of the weights each beneficiary receives in the model, the way that the effect of covariates is allowed to vary over calendar time and over time since the covariate values were measured, and how the regressions account for correlation in outcomes across quarters for a given individual.

g. Primary tests

Table II.C.1 shows the primary tests for Denver Health, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important, expressed as a percentage of change from the counterfactual—that is, the outcome the treatment group would have had in the absence of the HCIA-funded intervention. The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for detail and for a description of how we selected each test).

Table II.C.1. Specification of the primary tests for Denver Health

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population[c]	Substantive threshold (impact as percentage of comparison group mean)[d]
Quality-of-care outcomes (0)	n.a.—Awardee did not explicitly plan to affect the quality-of-care outcomes we can measure in claims	n.a.	n.a.	n.a.
Service use (4)	All-cause hospital admissions (#/person/quarter)	Intervention quarters 5 through 11	All Medicare FFS and Medicaid FFS	-5.0%
	All-cause outpatient ED visits (#/person/quarter)			-5.0%
	All-cause hospital admissions (#/person/quarter)		Medicare or Medicaid FFS high-risk beneficiaries	-15.0%
	All-cause outpatient ED visits (#/person/quarter)			-15.0%
Spending (4)	Total Medicare Part A and B spending and Medicaid FFS spending ($/person/month)	Intervention quarters 5 through 11	All Medicare FFS and Medicaid FFS	-2.4%
	Total Medicare Part A and B spending and Medicaid FFS spending ($/person/month)		Medicare or Medicaid FFS high-risk beneficiaries	-15.0%
	Total Medicare Part A and B spending and Medicaid FFS spending ($/person/month)	Intervention quarters 8 through 11	All Medicare FFS and Medicaid FFS	-2.6%
	Total Medicare Part A and B spending and Medicaid FFS spending ($/person/month)		Medicare or Medicaid FFS high-risk beneficiaries	-15.0%

Notes: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

High-risk beneficiaries are defined as follows: (1) beneficiaries who meet the program's eligibility criteria related to frequent acute care use—that is, those with three or more acute care visits (hospital admissions, ED visits, or urgent care visits) in the 12 months before joining the treatment group, or two or more such visits and a serious mental health diagnosis; and (2) (Medicaid only) any beneficiaries 18 and younger who meet Denver Health's diagnostic eligibility criteria for children with special health care needs.

[a] We adjusted the *p*-values from the primary test results for multiple comparisons made within each domain, but not across domains.

[b] The regressions we used for preliminary impact analysis have controlled for differences between the treatment and comparison groups before assignment to the analytic sample, as well as differences in outcomes in the baseline period.

[c] To specify the primary tests, we have made assumptions about the Medicaid data that will be available by the scheduled end of the evaluation in 2017. We believe it is possible we could have Medicaid data for Colorado covering the period through mid-summer 2015 (intervention quarter 11). However, this assumption is optimistic. In the event that Medicaid data are available for only a shorter period, we will revise our primary test timing to cover, for the Medicare population, quarters 5 through 11 and, for the Medicaid population, as many quarters as we are able to include within this same period.

[d] For total Medicare FFS and Medicaid FFS spending among the full population, we set the substantive threshold to 75 percent of Denver Health's expected effect in each relevant time period. For the other outcomes or subpopulations, for which Denver Health did not set an explicit target, we set the threshold equal to reductions in acute-care use or spending that Peikes et al. (2011) indicated could be feasible among either high-risk or general-population beneficiaries (as applicable) in a patient-centered medical home program.

ED = emergency department; FFS = fee-for-service.

n.a. = not applicable.

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** Denver Health's central goal was to reduce spending, so our impact analyses to date have assessed effects on Medicare Part A and B spending. In future reports, we plan to add total Medicaid FFS spending as well, pending available data. Because the awardee expects to achieve its cost reductions through decreased use of acute care services, we have also calculated preliminary results for the impacts on all-cause admissions and ED visits.
- **Time period.** Denver Health expected impacts to grow over time, with small impacts during the first year of the program and more substantial impacts in the second and third years of the program. Most of our primary tests thus cover both the second and third years of 21st Century Care—that is, the period starting in I5, which began on November 1, 2013. In the spending domain, however, we chose to analyze impacts both in the final year of the program's operation (that is, I8 through I11) as well as in the combined seven quarters fromthe start of the program's second year (I5 through I11). Analyzing impacts over the longer time period (seven quarters rather than four) allows for more stable estimates, based on more data, but with a smaller anticipated effect size.
- **Population.** Denver Health's impacts should be concentrated among beneficiaries who receive intensive services, but this population is small compared with the full population served by 21st Century Care. Because there are trade-offs between analyzing the high-risk subpopulation (for which expected effects are large but the sample size is moderate) and analyzing the entire Medicare FFS population (which is more representative of the program population served, but with smaller anticipated effects), we assess both in our primary tests. For this report, we used the definition of *high-risk beneficiaries* provided in Section II.C.2.b. If future reports include Medicaid data, we will expand the treatment group definition to include Medicaid and Medicare FFS beneficiaries. Specifically, we will include Medicaid FFS beneficiaries who qualify for the Denver Health program population due to frequent use of acute care services and, in addition, all beneficiaries 18 and younger who meet Denver Health's diagnostic criteria for CSHCN.
- **Direction (sign) of the impact estimate.** The primary tests test for a *reduction,* relative to the counterfactual, for each of the three outcome measures.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we specified thresholds for what we call substantive importance. Denver Health stated its anticipated effects on spending for the full population (Section I), so our substantive threshold for Medicare spending (or Medicaid spending, if applicable) is based on the awardee's value. We set the thresholds for substantive importance at 75 percent of the awardee's expected savings, which were 3.4 percent in the third year and 3.2 percent in the second and third years combined. (We used 75 percent of the awardee's anticipated value recognizing that Denver Health could still be considered successful if it approached, but did not achieve, its fully anticipated effects). Because the awardee did not specify anticipated impacts on intermediate outcomes or among subpopulations, we took all of our other thresholds—for spending among a high-risk population, for outpatient ED visits among either a high-risk or full population, or for all-cause admissions among either a high-risk or

full population—from the literature (Peikes et al. 2011). These thresholds are based on the assumption that a successful primary care intervention could cause a reduction in spending or service use of 5 percent among a general population and 15 percent among a high-risk population (Peikes et al. 2011).

As noted earlier, due to limitations in data availability, we were able to conduct the primary tests in this report only partially. Specifically, we estimated regression-adjusted differences only during I5 through I9 because we do not yet have data for I10 or I11 (February through July 2015). Our third annual report will cover the full period from I5 through I11. We might also add Medicaid beneficiaries, if data are available.

h. Secondary tests

We conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the non-experimental design of our study or random fluctuations in the data. We will have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. Specifically, we conducted three sets of secondary tests.

First, we tested whether the trends in outcomes for the treatment and comparison groups were parallel during the baseline period. Baseline outcomes were outcomes observed after matching (so they are not guaranteed to be similar between the treatment and comparison groups) but before the 21st Century Care program began (so differences between the treatment and comparison groups could not be caused by the HCIA-funded intervention). The assumption of parallel trends in the baseline period is a key assumption for the difference-in-differences model. If the trends are not parallel during the baseline period, it is difficult to interpret the baseline differences between the treatment and comparison groups as stable differences that would persist over time in the absence of the intervention—in short, differences that can be subtracted out during the intervention period to yield estimates of the program impacts.

Second, we estimated the program's impacts on ED visits, hospitalizations, and total Medicare Part A and B spending during the first year of the program: that is, in I1 through I4. Because Denver Health expected 21st Century Care to have only very small impacts during this first year, these tests provide a benchmark against which to judge the primary test results. If we were to find large impact estimates during the first year of the program, before they were expected, this could suggest a limitation in the comparison group or some other aspect of the study design, rather than true program impacts.

Third, we repeated the primary tests, but limited the sample to people who did not receive services from a clinic participating in CMMI's FQHC Demonstration at any time during the baseline or intervention periods. Because three of Denver Health's community health centers were participating in the FQHC Demonstration before the HCIA program began, the effects of the FQHC Demonstration could confound our estimates of 21st Century Care effectiveness. We have controlled for beneficiaries' contact with the FQHC Demonstration in our regression analyses (controlling for whether beneficiaries ever visited a Demonstration health center), but

because of the high proportion of Denver Health beneficiaries affected by the FQHC Demonstration (25 percent), we were not able to match the treatment and comparison groups well on this variable (Section II.C.4). These secondary tests, restricting analysis to the non-Demonstration population, help us to assess whether impact estimates look different for the non-Demonstration members of the treatment group than the full treatment group. This, in turn, provides information to assess whether our regression adjustment is sufficient to deal with any potential influence of the FQHC Demonstration on our estimates of the marginal impact of the HCIA program.

i. Synthesizing evidence to draw conclusions

We have developed a framework for drawing conclusions about program effectiveness within each domain, based on the primary test results, the results of secondary tests, and the plausibility of those findings given the implementation evidence. The four possible conclusions we can draw using this framework are as follows: (1) statistically significant favorable effect (the highest level of evidence), (2) substantively important favorable effect, (3) indeterminate effect, and (4) substantively important unfavorable effect. We cannot conclude that a program has a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program harms.

Our decision rules for each of the four possible conclusions are described in detail in Appendix 3. In short, we conclude that a program had a statistically significant favorable effect in a domain if (1) at least one primary test result in the domain is favorable and statistically significant, after adjusting the statistical tests to account for multiple tests within a domain; or (2) the average impact estimate across all primary tests in the domain is favorable and statistically significant. In both cases, we also have to determine that the primary test results are plausible given the secondary tests and implementation evidence. We conclude that a program had a substantively important favorable effect if the average impact estimate is substantively important but not statistically significant, and if the result is plausible given the secondary tests and implementation evidence. In contrast, if the average impact estimate is unfavorable (opposite the hypothesized direction) and larger than the substantive threshold, and unfavorable effects are plausible given the other evidence, we conclude the program had a substantively important unfavorable effect. Finally, if the tests in a domain do not meet any of these criteria, we conclude that the impact in that domain was indeterminate.

3. Characteristics of treatment beneficiaries when first assigned to the treatment group

This section describes the characteristics of the treatment group members at the time they were first assigned to the treatment group, which occurred from May 1, 2011, to June 30, 2014. We show these characteristics separately for the full population (in the second column of Table II.C.2.a) and for the high-risk subgroup (in the second column of Table II.C.2.b) because we present primary test results for both groups. (Tables II.C.2.a and II.C.2.b also serve a second purpose—to show the pre-intervention equivalence of the treatment and comparison groups—which we describe in Section II.C.4.)

Table II.C.2.a. Characteristics of the treatment and comparison group members when first assigned to their respective groups: All Medicare FFS beneficiaries

Beneficiary characteristic	Treatment group (N = 6,199)	Unmatched comparison group (N = 322,104)	Matched comparison group (N = 23,196)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Exact match variables[c]						
Program eligibility as a high-risk frequent user of acute care services (%)	26.4	11.5	26.4	0.00	0.00	n.a.
Caliper-matched variables[d]						
Age (years)	58.8	71.1	58.8	0.01	0.00	71[e]
Propensity-matched variables[f]						
Demographic characteristics						
Male (%)	50.5	45.0	51.2	-0.66	-0.01	45.3[g]
Dual status (%)	71.0	16.1	71.2	-0.29	-0.01	21.7[h]
Race: Black (%)	23.8	2.3	21.1	2.75***	0.07	10.4[g]
Race: Hispanic (%)	17.6	2.8	19.1	-1.53***	-0.04	2.6[g]
Zip code poverty rate (%)	20.9	12.8	21.0	-0.07	-0.01	NA
Zip code unemployment rate (%)	9.9	8.2	10.0	-0.05	-0.02	NA
Zip code percentage with college degree (%)	33.8	34.0	32.2	1.65***	0.09	NA
Original reason for Medicare entitlement (%)						
Disability	68.9	21.7	72.7	-3.79***	-0.08	16.7[g]
ESRD	0.8	0.3	0.6	0.15	0.02	0.1[g]
Both disability and ESRD	1.5	0.5	1.9	-0.38*	-0.03	NA
Health status and chronic conditions						
HCC risk score	1.3	1.1	1.3	0.02	0.02	1.0
Serious mental health diagnosis (%)	18.9	4.0	18.1	0.87	0.02	NA
Service use and spending 3 months before assignment						
Number of hospital admissions (#/1,000/quarter)	167.9	86.2	163.3	4.61	0.01	74[i]
Number of ED visits (#/1,000/quarter)	477.8	158.6	435.4	42.39**	0.03	105[j]
Medicare Part A and B spending ($/month)	1,436	837	1,441	-5.38	0.00	860[k]
Service use and spending 4 to 12 months before assignment						
Number of hospital admissions (#/1,000/quarter)	111.1	64.0	102.6	8.49*	0.03	74[i]
Number of ED visits (#/1,000/quarter)	340.8	126.6	325.2	15.59	0.02	105[j]
Medicare Part A and B spending ($/month)	1,060	484	984	76.23	0.03	860[k]
Other service use						
Number of unplanned readmissions (#/1,000/quarter in the 12 months before assignment)	28.0	8.7	23.3	4.75*	0.03	NA
Number of primary care visits (#/1,000/quarter in the 18 months before assignment)	740.0	972.1	790.2	-50.22***	-0.05	NA
Variables not matched						
FQHC Demonstration participation[l] (%)	25.4	1.6	6.9	18.48***	0.64	NA
Omnibus test for balance on matching variables[m]						
p-value			<0.001			

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services. Zip code data merged from the American Community Survey ZIP Code Characteristics.

Table II.C.2.a *(continued)*

Notes: The matched comparison group means are weighted based on the number of matched comparisons per treatment beneficiary. For example, if four comparison beneficiaries are matched to one treatment beneficiary, each of the four comparison beneficiaries has a matching weight of 0.25. To calculate the *unmatched* comparison group means we used one (unweighted) observation per beneficiary. For the beneficiaries in this pool who did eventually match we used the person-month observation from the month in which they matched. For those who did not match to a treatment beneficiary we used the first eligible person-month observation (from May 2011 to June 2014).

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the matched treatment and comparison groups. A positive number reflects a higher value among the treatment group on the matching variable and a negative number reflects a higher value among the comparison group.

[b] The standardized difference is the difference in means between the matched treatment and comparison groups divided by the standard deviation of the variable; the standard deviation is calculated across the pooled treatment and matched comparison groups.

[c] Exact match means that a beneficiary qualifying for the 21st Century Care program population as a frequent user of EDs, hospitals, or urgent care clinics could match only to another beneficiary who met the same criteria. (We used these same criteria to define the high-risk subgroup.) A beneficiary qualifying for the program population by meeting other criteria (that is, through primary care use only) could match only to other beneficiaries not meeting the high-risk criteria.

[d] Caliper match means we forced each treatment beneficiary to have a value for a given variable that fell within a specified range. We caliper-matched on age, requiring that each treatment beneficiary match to comparison beneficiaries no more than five years older or younger.

[e] Health Indicators Warehouse (2014a).

[f] Variables that we matched on through a propensity score, which captures the relationship between a beneficiary's characteristics and his or her likelihood of being in the treatment group.

[g] Chronic Conditions Warehouse (2014a, Table A1).

[h] Health Indicators Warehouse (2014c).

[i] Health Indicators Warehouse (2014b).

[j] Gerhardt et al. (2014).

[k] Boards of Trustees (2013).

[l] Beneficiaries are considered to be FQHC Demonstration participants if they visited an FQHC participating in the Demonstration at any time from November 2011 (when the Demonstration started) to June 2014.

[m] Results from an overall chi-squared test indicate the likelihood of observing differences in the matching variables as large as the differences we observed if, in fact, the treatment and comparison populations (from which we drew the samples) were perfectly balanced. The value of $p < 0.001$ for the chi-squared test suggests there is an extremely small chance that populations are identical, even though the magnitude of all differences is small. The omnibus test does not include the variable for FQHC Demonstration participation, as this was not a matching variable.

*/**/*** Significantly different from zero at the .10/.05/.01 levels, respectively, two-tailed test. (Note: The primary tests assume a one-tailed tests, for the reasons explained in the text.)

ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; FQHC = Federally Qualified Health Center; HCC = Hierarchical Condition Category.

NA = not available.

n.a. = not applicable.

Table II.C.2.b. Characteristics of the treatment and comparison group members when first assigned to their respective groups: High-risk Medicare FFS beneficiaries

Beneficiary characteristic	Treatment group (N = 1,636)	Unmatched comparison group (N = 36,992)	Matched comparison group (N = 6,526)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Caliper matched variables[d]						
Age (years)	54.7	68.6	54.7	-0.02	0.00	71[e]
Propensity-matched variables[f]						
Demographic characteristics						
Male (%)	55.0	44.3	55.0	-0.05	0.00	45.3[g]
Dual Medicare/Medicaid status (%)	74.5	35.3	75.5	-1.04	-0.02	21.7[h]
Race: Black (%)	22.6	4.0	19.0	3.51***	0.10	10.4[g]
Race: Hispanic (%)	13.3	4.7	13.7	-0.43	-0.01	2.6[g]
Zip code poverty rate (%)	20.1	14.2	20.1	0.06	0.01	NA
Zip code unemployment rate (%)	9.7	8.6	9.8	-0.11	-0.03	NA
Zip code proportion with college degree (%)	35.1	31.7	33.5	1.67***	0.09	NA
Original reason for entitlement (%)						
Disability	77.3	40.2	82.4	-5.14***	-0.13	16.7[g]
ESRD	1.3	1.0	0.8	0.56*	0.06	0.1[g]
Both disability and ESRD	2.4	1.5	3.3	-0.91*	-0.06	NA
Health status and chronic conditions						
HCC risk score	2.4	2.4	2.3	0.05	0.03	1.0
Serious mental health diagnosis (%)	34.2	15.0	33.4	0.82	0.02	NA
Service use and spending 3 months before assignment						
Number of hospital admissions (#/1,000/quarter)	507.3	437.4	480.8	26.54	0.03	74[i]
Number of ED visits (#/1,000/quarter)	1,561.1	868.1	1,395.1	166.02**	0.08	105[j]
Medicare spending ($/month)	3,901	3,406	3,883	17.92	0.00	860[k]
Service use and spending 4 to 12 months before assignment						
Number of hospital admissions (#/1,000/quarter)	334.8	321.6	302.1	32.65**	0.07	74[i]
Number of ED visits (#/1,000/quarter)	1,072.3	682.2	1,006.4	65.96	0.05	105[j]
Medicare Part A and B spending ($/month)	2,746	1,982	2,528	218.43*	0.05	860[k]
Other service use						
Number of unplanned readmissions (#/1,000/quarter in the 12 months before assignment)	103.5	68.2	86.0	17.50*	0.06	NA
Number of primary care visits (#/1,000/quarter in the 18 months before assignment)	1,185	1,517	1,251	-66.08*	-0.05	NA
Variables not matched						
FQHC Demonstration participation[l] (%)	17.2	2.8	7.8	9.4***	0.33	
Omnibus test for balance on matching variables[m]						
p-value			<0.001			

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services. Zip code household income data merged from the American Community Survey ZIP Code Characteristics.

Table II.C.2.b *(continued)*

Notes: The matched comparison group means are weighted based on the number of matched comparisons per treatment beneficiary. For example, if four comparison beneficiaries are matched to one treatment beneficiary, each of the four comparison beneficiaries has a matching weight of 0.25. To calculate the *unmatched* comparison group means we used one (unweighted) observation per beneficiary. For the beneficiaries in this pool who did eventually match we used the person-month observation from the month in which they matched. For those who did not match to a treatment beneficiary we used the first eligible person-month observation (from May 2011 to June 2014).

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the matched treatment and comparison groups. A positive number reflects a higher value among the treatment group on the matching variable and a negative number reflects a higher value among the comparison group.

[b] The standardized difference is the difference in means between the matched treatment and comparison groups divided by the standard deviation of the variable; the standard deviation is calculated across the pooled treatment and matched comparison groups.

[c] Exact match means that a beneficiary qualifying for the 21st Century Care program population as a frequent user of EDs, hospitals, or urgent care clinics could match only to another beneficiary who met the same criteria. (We used these same criteria to define the high-risk subgroup.) A beneficiary qualifying for the program population by meeting other criteria (that is, through primary care use only) could match only to other beneficiaries not meeting the high-risk criteria.

[d] Caliper match means we forced each treatment beneficiary to have a value for a given variable that fell within a specified range. We caliper-matched on age, requiring that each treatment beneficiary match to comparison beneficiaries no more than five years older or younger.

[e] Health Indicators Warehouse (2014a).

[f] Variables that we matched on through a propensity score, which captures the relationship between a beneficiary's characteristics and his or her likelihood of being in the treatment group.

[g] Chronic Conditions Warehouse (2014a, Table A1).

[h] Health Indicators Warehouse (2014c).

[i] Health Indicators Warehouse (2014b).

[j] Gerhardt et al. (2014).

[k] Boards of Trustees (2013).

[l] Beneficiaries are considered to be FQHC Demonstration participants if they visited an FQHC participating in the Demonstration at any time from November 2011 (when the Demonstration started) to June 2014.

[m] Results from an overall chi-squared test indicate the likelihood of observing differences in the matching variables as large as the differences we observed if, in fact, the treatment and comparison populations (from which we drew the samples) were perfectly balanced. The value of $p < 0.001$ for the chi-squared test suggests there is an extremely small chance that populations are identical, even though the magnitude of all differences is small. The omnibus test does not include the variable for FQHC Demonstration participation, as this was not a matching variable.

*/**/*** Significantly different from zero at the .10/.05/.01 levels, respectively, two-tailed test. (Note: The primary tests assume a one-tailed tests, for the reasons explained in the text.)

ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; FQHC = Federally Qualified Health Center; HCC = Hierarchical Condition Category.

NA = not available.

n.a. = not applicable.

Characteristics of the treatment group members overall. Consistent with Denver Health's role as a safety net provider, the treatment beneficiaries were unusually likely to have gained Medicare FFS eligibility for a reason other than old age (with 69 percent qualifying due to disability, compared with 17 percent nationwide). As a result, the treatment group was relatively young (with a mean age of 59 years, compared with a Medicare FFS national average of 71). More than two-thirds (71 percent) of the treatment beneficiaries were dually eligible for both Medicare and Medicaid.

The treatment group members also had relatively high service use and Medicare FFS spending in the year before assignment, compared with the national averages. For example, the ED visit rate in the three months directly before treatment group assignment (478 per 1,000 beneficiaries per quarter) was more than four times the FFS national average. The admissions rate (168 per 1,000 beneficiaries per quarter) was more than double and spending ($1,436 per person per month) was 67 percent higher than the national average over the same period. This is consistent both with Denver Health serving a generally high-needs population and with the awardee's definition of its program population to include not only primary care users but also members of the high-risk subgroup, whose characteristics we describe next.

Characteristics of the high-risk treatment group members. The high-risk population is defined as beneficiaries with three or more acute care visits (hospitalizations, ED visits, or urgent care visits) in the past 12 months or with two or more hospitalizations in the past 12 months and a qualifying serious mental health diagnosis. Unsurprisingly, then, we see that the high-risk subgroup had very high rates of hospitalization and ED visits in the months leading up to assignment. For example, in the three months before assignment, the ED visit rate (1,072 per 1,000 beneficiaries per quarter) was more than 10 times the national average, and the hospital admission rate (335 per 1,000 beneficiaries per quarter) was more than 4 times the national average. Spending over the same period ($2,746 per person per month) was more than 3 times the national average. Eighty-one percent of the beneficiaries were eligible for Medicare for a reason other than old age and 74 percent were dually eligible for Medicare and Medicaid.

4. Equivalence of the treatment and comparison groups at the time of matching

Demonstrating that the treatment and comparison groups were similar before beneficiaries began receiving HCIA-funded services is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying difference-in-differences models—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment beneficiaries not received the intervention.

Equivalence of the full treatment and comparison groups. Table II.C.2.a shows that the 6,199 treatment beneficiaries and the 23,196 selected comparison beneficiaries were similar when they were assigned to their respective groups (treatment or comparison), before the treatment beneficiaries could have received 21st Century Care services. By construction, there were no differences between the two groups on the exact matching variable—whether the beneficiary had sufficient service use to qualify for the high-risk subgroup. Although there were some differences between treatment beneficiaries and matched comparison group beneficiaries on the variables we matched through caliper matching or propensity scores, these differences

were very small. The standardized differences across the matching variables were all within 0.10 and most were within 0.05—well within our matching target of 0.25 standardized differences. (The 0.25 target is an industry standard; for example, see Institute of Education Sciences [2014]). This suggests that any residual differences on the matching variables between the treatment and comparison groups can be accounted for in the impact analysis through regression adjustment.

Matching improved our balance on almost all variables relative to the unmatched comparison pool column three of Table II.C.2.a). This improvement was especially dramatic for Medicare/Medicaid dual eligibility status, age, black race, Hispanic ethnicity, proportion originally entitled to Medicare due to disability, and all of the service use variables. This is not surprising given that the comparison pool included every Medicare FFS beneficiary in Colorado outside of Denver who had had a primary care visit in the previous 18 months. Matching made the equivalence between treatment and comparison groups worse for only one variable: proportion in the beneficiary's zip code with a college degree. As noted previously, however, even this difference is well within our target of 0.25 standardized differences.

Of the differences on the matching variables that remained after matching, the largest standardized differences were for black race (24 percent prevalence among the treatment group compared with 21 percent among the comparison), proportion in the zip code with a college degree (34 percent prevalence compared with 32 percent), and proportion originally entitled to Medicare due to disability (69 percent compared with 73 percent). Although the differences on the matching variables are small, many are still statistically significant because of the large sample sizes involved. Based on an omnibus test, we can soundly reject the hypothesis that the groups are identical across all matching variables ($p < 0.001$). This underscores the importance of regression-adjusting for any residual differences after matching in any of the matching variables, as we have done in our analysis. Statistically significant differences are not in themselves a sign that the groups are poorly matched when the differences are small in magnitude.

As noted earlier, we were not able match on receiving services from a health center participating in the FQHC Demonstration, due to the small number of potential comparison beneficiaries who visited such health centers. About 25 percent of the treatment beneficiaries received services at some point from a community health center participating in the FQHC Demonstration, compared with only 7 percent of the comparison beneficiaries (standardized difference = 0.64). We have regression-adjusted for this variable in our impact analyses and, as described previously, have conducted secondary tests to assess the risk of possible bias from inability to match on this variable directly.

Equivalence for the high-risk subgroup. Table II.C.2.b shows that the 1,636 high-risk treatment beneficiaries were similar to the 6,526 high-risk comparison beneficiaries when they entered the analytic sample. As with the full treatment group, the standardized differences across all matching variables are well within our target of 0.25 standardized differences. The largest standardized difference (for the proportion originally entitled to Medicare due to disability) is 0.13. All other differences are within 0.10 standardized differences and most are 0.05 or less. As with the overall sample, matching improved our balance relative to the unmatched comparison

pool for almost all variables, even though in this case the pool was substantially narrower (and thus more similar to the treatment group), defined to be people meeting the high-risk criteria of three acute care visits in a 12-month period or two hospitalizations with a qualifying mental health diagnosis. Matching improved balance on all variables except for HCC score (for which the difference was unchanged before and after matching) and for the proportion originally entitled to Medicare due to ESRD and the number of hospitalizations 4 to 12 months before assignment to the treatment or comparison group (for which differences were marginally worse after matching).

As with the sample overall, many differences on the matching variables are still statistically significant because of the large sample sizes. Based on the omnibus test, we can reject the hypothesis that the groups are identical on all matching variables ($p < 0.001$), even though the differences are small. This again underscores the importance of regression-adjustment to account in the impact analysis for residual differences in the matching variables. Also, as with the sample overall, we did not match on FQHC Demonstration participation (17 percent among the high-risk treatment beneficiaries and 8 percent among their matched comparison counterparts; standardized difference = 0.33). We have regression-adjusted for this variable and assessed the possible effects of the lack of equivalence through our secondary tests.

5. Intervention impacts

In this section, we present the preliminary results of our impact analysis. We first present sample sizes and mean outcomes, by quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates; however, the differences in mean outcomes are not impact estimates by themselves. Next, we present the results of the primary tests (which are regression-adjusted), by domain. Then, we present the secondary tests results and assess the plausibility of the primary test results given the secondary tests and the implementation evidence. We then show how the preliminary evidence gathered so far fits into our framework for drawing conclusions, and end with a discussion of next steps for the impact evaluation.

a. Sample sizes

We present sample sizes separately for all beneficiaries (Table II.C.3.a) and for the high-risk subgroup (Table II.C.3.b) because our primary tests include both populations. Over the six baseline quarters. The number of assigned beneficiaries among the full population decreased from 3,634 to 3,488 in the treatment group and from 12,122 to 11,836 in the comparison group. Over the nine intervention quarters, the numbers then decreased further from 2,916 to 1,690 in the treatment group and from 10,105 to 5,434 in the comparison group. Among the high-risk subgroup, the number of assigned beneficiaries in the baseline period increased slightly from 694 to 714 among the treatment group and from 2,344 to 2,550 among the comparison group, and then decreased during the intervention period from 534 to 354 among the treatment group and from 1,954 to 1,177 among the comparison group.

Table II.C.3.a. Unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and comparison groups for Denver Health, by quarter: May 1, 2011, to January 31, 2015

	Number of Medicare FFS beneficiaries			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/month)		
Quarter	T	C (no wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (May 1, 2011 – October 31, 2012)												
B1	3,634	11,727	3,634	122.5	96.2	26.2 (27.2%)	283.7	278.3	5.4 (1.9%)	$1,137	$982	$154 (15.7%)
B2	3,680	12,122	3,680	105.2	88.5	16.7 (18.9%)	285.7	270.2	15.5 (5.7%)	$988	$965	$22 (2.3%)
B3	3,665	12,286	3,665	104.0	83.0	21.0 (25.3%)	251.4	263.0	-11.7 (-4.4%)	$1,022	$923	$100 (10.8%)
B4	3,349	11,061	3,349	104.5	84.3	20.2 (24.0%)	299.5	278.2	21.3 (7.7%)	$1,052	$919	$133 (14.5%)
B5	3,405	11,395	3,405	98.1	77.8	20.3 (26.1%)	355.0	295.6	59.5 (20.1%)	$1,034	$858	$176 (20.5%)
B6	3,488	11,836	3,488	94.3	90.1	4.3 (4.7%)	335.7	309.9	25.8 (8.3%)	$1,043	$954	$89 (9.3%)
Intervention period (November 1, 2012 – January 31, 2015)												
I1	2,916	10,105	2,916	107.3	69.9	37.5 (53.6%)	340.7	264.4	76.3 (28.9%)	$1,039	$852	$187 (21.9%)
I2	2,595	8,784	2,595	108.7	83.2	25.4 (30.6%)	330.8	273.8	57.0 (20.8%)	$980	$909	$71 (7.8%)
I3	2,527	8,505	2,527	116.7	74.5	42.3 (56.8%)	402.6	322.7	79.9 (24.8%)	$1,159	$872	$287 (33.0%)
I4	2,458	8,315	2,458	115.5	80.1	35.4 (44.2%)	396.3	297.3	99.0 (33.3%)	$1,067	$944	$123 (13.0%)
I5	2,368	8,006	2,368	111.9	67.5	44.4 (65.8%)	409.2	287.0	122.2 (42.6%)	$1,157	$837	$320 (38.2%)
I6	2,110	6,951	2,110	101.9	78.8	23.1 (29.3%)	422.7	301.4	121.3 (40.2%)	$1,093	$892	$201 (22.5%)
I7	1,985	6,566	1,985	114.4	74.5	39.9 (53.6%)	411.5	313.0	98.5 (31.5%)	$1,165	$899	$265 (29.5%)
I8	1,843	6,062	1,843	111.2	76.4	34.8 (45.5%)	446.2	340.1	106.1 (31.2%)	$1,119	$947	$171 (18.1%)
I9	1,690	5,434	1,690	99.4	84.5	14.9 (17.6%)	435.1	283.8	151.3 (53.3%)	$1,218	$884	$335 (37.9%)

Table II.C.3.a *(continued)*

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The baseline quarters are measured relative to when the baseline period began on May 1, 2011. For example, the first baseline quarter (B1) runs from May 1, 2011, to July 31, 2011. The intervention quarters are measured relative to the start of the intervention period on November 1, 2012. For example, the first intervention quarter (I1) runs from November 1, 2012, to January 31, 2013.

In each period (baseline or intervention), the treatment group each quarter consists of Medicare FFS beneficiaries who had been, by the first date of the quarter, (1) attributed to Denver Health's program population and (2) continuously observable for the previous 12 months in FFS claims data. (Outcomes are observable if the beneficiary is alive, enrolled in FFS Medicare [Part A and B], and has Medicare as the primary payer.) In the intervention period, we further excluded beneficiaries from the treatment group if Denver Health considered them part of the program population before they had 12 months of Medicare FFS claims observability (that is, before we could match them to comparison beneficiaries). In each quarter, the comparison group consists of Medicare FFS beneficiaries matched to the treatment beneficiaries. The means presented in this table are for complete matched sets only. That is, if either the treatment group beneficiary or *any* of the matched comparison group members in a matched set are not observable in a quarter, any remaining beneficiaries in the matched set are removed from the sample in that quarter.

The outcome means were weighted, such that (1) each treatment beneficiary gets a weight of 1 and (2) each comparison beneficiary gets a weight that is equal to the reciprocal of the total number of comparison beneficiaries matched to the treatment beneficiary. For example, if a treatment beneficiary was matched to four comparison beneficiaries, each comparison beneficiary receives a weight of 0.25. The difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; Diff = difference; ED = emergency department; FFS = fee-for-service; I = intervention; T = treatment; wgt = weight.

Table II.C.3.b. Unadjusted mean outcomes for high-risk Medicare FFS beneficiaries in the treatment and comparison groups for Denver Health, by quarter: May 1, 2011, to January 31, 2015

	Number of Medicare FFS beneficiaries			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/month)		
Quarter	T	C (no wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (May 1, 2011 – October 31, 2012)												
B1	694	2,344	694	357.3	276.1	81.2 (29.4%)	883.3	927.7	-44.4 (-4.8%)	$3,243	$2,389	$854 (35.7%)
B2	691	2,395	691	314.0	231.9	82.2 (35.4%)	946.5	815.6	130.9 (16.0%)	$2,706	$2,331	$375 (16.1%)
B3	688	2,439	688	287.8	194.5	93.3 (48.0%)	829.5	825.7	3.8 (0.5%)	$2,306	$2,120	$186 (8.8%)
B4	644	2,280	644	274.8	201.4	73.4 (36.5%)	987.6	868.8	118.8 (13.7%)	$2,218	$1,833	$385 (21.0%)
B5	673	2,428	673	227.3	192.8	34.5 (17.9%)	1,135.9	862.4	273.5 (31.7%)	$2,006	$1,783	$223 (12.5%)
B6	714	2,550	714	259.1	209.9	49.2 (23.5%)	1,080.4	871.6	208.8 (24.0%)	$2,193	$1,945	$248 (12.8%)
Intervention period (November 1, 2012 – January 31, 2015)												
I1	534	1,954	534	299.6	176.8	122.8 (69.5%)	1,119.5	802.8	316.7 (39.4%)	$2,327	$1,805	$521 (28.9%)
I2	472	1,673	472	341.1	201.1	140.0 (69.6%)	1,166.3	827.6	338.7 (40.9%)	$2,425	$1,966	$459 (23.3%)
I3	468	1,644	468	329.1	195.7	133.4 (68.2%)	1,369.5	963.3	406.3 (42.2%)	$2,770	$1,971	$798 (40.5%)
I4	476	1,686	476	344.5	179.4	165.2 (92.1%)	1,327.7	808.7	519.0 (64.2%)	$2,699	$1,784	$915 (51.3%)
I5	473	1,664	473	296.0	150.6	145.4 (96.6%)	1,419.1	798.0	621.1 (77.8%)	$2,606	$1,671	$935 (56.0%)
I6	451	1,578	451	261.6	173.3	88.4 (51.0%)	1,310.0	833.5	476.5 (57.2%)	$2,469	$1,696	$772 (45.5%)
I7	427	1,489	427	304.4	172.6	131.9 (76.4%)	1,191.9	865.3	326.6 (37.7%)	$2,584	$1,855	$728 (39.3%)
I8	408	1,424	408	286.8	156.6	130.1 (83.1%)	1,227.7	894.4	333.3 (37.3%)	$2,519	$1,840	$679 (36.9%)
I9	354	1,177	354	211.9	183.9	27.9 (15.2%)	1,255.2	712.0	543.2 (76.3%)	$2,316	$1,734	$582 (33.5%)

Table II.C.3.b *(continued)*

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The baseline quarters are measured relative to when the baseline period began on May 1, 2011. For example, the first baseline quarter (B1) runs from May 1, 2011, to July 31, 2011. The intervention quarters are measured relative to the start of the intervention period on November 1, 2012. For example, the first intervention quarter (I1) runs from November 1, 2012, to January 31, 2013.

In each period (baseline or intervention), the high-risk subgroup of the treatment group each quarter consists of Medicare FFS beneficiaries who had been, by the first date of the quarter, (1) attributed to Denver Health's program population due to frequent use of acute care services (with three or more hospital admissions, ED visits, or urgent care visits in the past 12 months, or two or more hospital admissions with a qualifying mental health diagnosis); and (2) continuously observable for the previous 12 months in FFS claims data. (Outcomes are observable if the beneficiary is alive, enrolled in FFS Medicare [Part A and B], and has Medicare as the primary payer.) In the intervention period, we further excluded beneficiaries from the high-risk subgroup of the treatment group if Denver Health considered them part of the program population before they had 12 observable months of Medicare FFS claims (that is, before we could match them to comparison beneficiaries). In each quarter, the high-risk subgroup of the comparison group consists of Medicare FFS beneficiaries matched to the high-risk treatment beneficiaries. The means presented in this table are for complete matched sets only. That is, if either the treatment group beneficiary or *any* of the matched comparison group members in a matched set are not observable in a quarter, any remaining beneficiaries in the matched set are removed from the sample in that quarter.

The outcome means were weighted, such that (1) each treatment beneficiary gets a weight of 1 and (2) each comparison beneficiary gets a weight that is equal to the reciprocal of the total number of comparison beneficiaries matched to the treatment beneficiary. For example, if a treatment beneficiary was matched to four comparison beneficiaries, each comparison beneficiary receives a weight of 0.25. The difference between the high-risk subgroups of the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the high-risk comparison subgroup from the mean outcome for the high-risk treatment subgroup. The percent difference equals that difference divided by the mean outcome for the comparison subgroup.

B = baseline; C = comparison; Diff = difference; ED = emergency department; FFS = fee-for-service; I = intervention; T = treatment; wgt = weight.

The observed decreases in sample size occur because we have limited the analytic sample to complete matched sets only; that is, we dropped all members of a matched set whenever any member of the matched set exited the sample (for example, due to death or to membership in a Medicare Advantage plan). A decreasing sample size thus reflects the fact that matched sets have exited the sample over time at a higher rate than new matched sets joined (by treatment group members meeting the Denver Health eligibility criteria with 12 months of Medicare FFS claims history). The decreases in sample size are more pronounced during the intervention than baseline period because, as explained in Section II.C.2.b, we assigned treatment beneficiaries during the intervention period only if they already had 12 months of claims history when they first met the Denver Health program eligibility criteria.

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

Admissions. Among the full population (Table II.C.3.a), the number of all-cause inpatient admissions among the treatment beneficiaries declined steadily during the baseline period from 122.5 to 94.3 per 1,000 beneficiaries per quarter. Admissions were consistently lower—by 4.3 to 26.2 percent—among the comparison beneficiaries during the same period, and also generally declined over time. This difference between the treatment and comparison group during the baseline period, before the intervention began (but after matching), could reflect either (1) a so-called facility effect—that is, a difference in the Denver Health environment relative to the comparison environment, such as an ability for urban dwellers in Denver to reach the hospital more easily than comparison beneficiaries in more rural areas; or (2) the effect of beneficiary characteristics that we could not match on because they are not observed in claims data, such as limited health literacy, limited English proficiency, or homelessness. Thus, even though the treatment and comparison groups were well matched on observable characteristics before entering the analytic sample, we observe differences in outcomes following matching, before the HCIA-funded program began. These differences during the baseline period motivate our analytic strategy of using a difference-in-differences model for estimating impacts during the intervention period.

During the intervention period, the number of admissions among the treatment group was roughly similar to the number during the baseline period, ranging from 99.4 to 116.7 per 1,000 beneficiaries per quarter, but with no obvious trend (that is, without the same obvious decline over time). Admissions among the comparison group were again lower than among the treatment group in every quarter. However, the differences between treatment and comparison groups widened during the intervention period, with the comparison group means in each quarter 17.6 to 65.8 percent lower than the treatment group means.

Among the high-risk subgroup (Table II.C.3.b), the number of all-cause inpatient admissions was higher than among the full population, as we expected given the population definition as people who frequently visit the hospital, the ED, or an urgent care clinic. Specifically, for the high-risk treatment beneficiaries, admissions ranged from 227.3 to 357.3 per 1,000 beneficiaries per quarter during the baseline period (generally declining over time) and from 211.9 to 344.5 during the intervention period (with no obvious trend). The number of admissions among the high-risk comparison beneficiaries was again lower than among the treatment beneficiaries: 17.9

to 48.0 percent lower in each quarter during the baseline period and 15.2 to 96.6 percent lower during the intervention period.

Outpatient ED visits. Among the full population, the ED visit rate among the treatment group increased considerably over time, ranging from 251.4 to 355.0 per 1,000 beneficiaries per quarter during the baseline period but reaching 446.2 per 1,000 beneficiaries per quarter in I8 and 435.1 in I9. The comparison group did not experience this same increase. Overall, during the baseline period, the ED visit rate for the comparison group was generally lower than for the treatment group, with means each quarter ranging from 4.8 percent *higher* to 31.7 percent lower than the treatment group means. However, during the intervention period the comparison group rates remained roughly constant while the treatment group rates increased, so that by I9 the difference between the treatment and comparison groups was 53.3 percent.

The pattern for outpatient ED visits was similar among the high-risk group, although, as expected, the total number of ED visits was much higher among the high-risk group than among the full population. The treatment group rate ranged from 829.5 to 1,135.9 per 1,000 beneficiaries quarter during the baseline period. In each quarter, the comparison group mean was 4.8 percent higher to 31.7 percent lower than the treatment group mean. During the intervention period, rates were generally higher for the treatment group than they had been during the baseline period, ranging from 1,119.5 to 1,419.1 per 1,000 beneficiaries per quarter. The rate among the high-risk comparison beneficiaries was 37.3 to 77.8 percent lower than among the high-risk treatment beneficiaries during the same period.

Spending. Among the full population, for the treatment group, total Medicare Part A and B spending ranged from $988 to $1,137 per beneficiary per month during the baseline period and from $980 to $1,218 during the intervention period. There was no obvious trend in treatment group spending over time. However, spending was higher in the treatment group than in the comparison group in every baseline and intervention quarter. Spending among the comparison group was 2.3 to 20.5 percent lower than among the treatment group during the baseline period and 7.8 to 38.2 percent lower during the intervention period.

Among the high-risk population, total Medicare Part A and B spending among treatment beneficiaries was $2,006 to $3,243 per beneficiary per month during the baseline period. Spending among the high-risk comparison beneficiaries was 8.8 to 35.7 percent lower during this period. During the intervention period, spending for the high-risk treatment beneficiaries was $2,316 to $2,770 per beneficiary per month, and spending among the high-risk comparison beneficiaries was 23.2 to 56.0 percent lower.

c. Results for primary tests

Overview. The primary tests are conducted using the quarterly data presented in Tables II.C.3.a and II.C.3.b, but the impact estimate for each outcome is an average over the specified time period for the test. As noted previously, based on available claims data, we conducted the primary tests in this report only for the period through January 2015: that is, either over the five intervention quarters from I5 to I9 (for the primary tests in the service use domain and for two of the four tests in the spending domain) or over I9 (for the remaining two tests in

the spending domain). Final results will cover the period through I11. All primary test results reported here are regression-adjusted for any differences that existed between the treatment and comparison groups during the baseline period. We also regression-adjusted for differences in beneficiaries' characteristics within each matched set and for matched-set-by-quarter fixed effects.

Tests in both domains indicate substantively important unfavorable effects of the 21st Century Care program. As described earlier, these results are preliminary and are presented here as a work in progress. Table II.C.4 presents the primary test results.

Service use. Among the full population, the 21st Century Care program was associated with a 7.3 percent increase in all-cause hospital admissions from I5 through I9 and an increase of 21.1 percent in the outpatient ED visit rate, relative to the estimated counterfactual. These increases correspond to an estimated 7.4 additional admissions and 73.9 additional ED visits per 1,000 beneficiaries per quarter. Among the high-risk subgroup, the program was associated with a 5.6 percent increase in admissions and a 17.7 percent increase in the outpatient ED visit rate, relative to the estimated counterfactual. This is equivalent to an additional 14.4 admissions and 192.4 additional ED visits per 1,000 beneficiaries per quarter. None of these differences is statistically significant because we used one-sided statistical tests to test for program impacts. (That is, we tested only for decreases in service use relative to the counterfactual, not the increases that we observe.) However, the mean effect estimate in the domain indicates an overall 12.9 percent increase in service use. This is unfavorable and larger than our threshold of 10 percent for substantive importance across all outcomes in the domain combined.

Table II.C.4 also shows that our tests had poor statistical power to detect substantively important favorable effects, if such effects existed. For example, if the program had succeeded in decreasing service use by 10 percent, we would have had only a 44.8 percent probability of finding a statistically significant difference in the mean effect estimate (using a one-tailed test, and a $p < 0.10$ threshold). This poor statistical power means that even relatively large impact estimates could be due to chance, rather than to true impacts.

Spending. The 21st Century Care program was associated with an 11.6 percent increase in total Medicare Part A and B spending among the full population from I5 through I9 and a 19.0 percent increase in I9 only (the first quarter of the final year of the program), relative to the estimated counterfactual. Among the high-risk population, the program was associated with an increase in spending of 9.2 percent from I5 through I9 and an increase of 7.7 percent in I9, relative to the counterfactual. These increases correspond to estimated additional expenditures of $120, $195, $211, and $166 per beneficiary per month, respectively.

As in the service use domain, none of these differences is statistically significant because we used one-sided statistical tests to test for reductions in spending. However, the mean effect estimate in the domain indicates an overall 11.9 percent increase in spending. This is unfavorable and larger than our combined threshold of 8.8 percent for substantive importance in the domain. Statistical power to detect a favorable effect the size of the substantive threshold is once again poor (26.4 percent), indicating that even relatively large impact estimates could be due to chance.

Table II.C.4. Results of primary tests for Denver Health

Primary test definition					Statistical power to detect an effect that is[a]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual)[b]	Size of the substantive threshold	Twice the size of the substantive threshold	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual (standard error)[a]	Percentage difference[c]	*p*-value[d]
Service use (4)	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 5–9	All observable Medicare FFS beneficiaries	-5.0%	24.6	46.3	107.8	7.4 (8.4)	7.3%	0.623[e]
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			-5.0%	23.0	42.2	424.9	73.9 (32.4)	21.1%	0.966[e]
	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)		All observable high-risk Medicare FFS beneficiaries	-15.0%	49.2	89.3	272.1	14.4 (30.6)	5.6%	0.522[e]
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			-15.0%	54.3	93.3	1,280.8	192.4 (117.5)	17.7%	0.865[e]
	Combined (%)			-10.0%	44.8	84.6	n.a.	n.a.	12.9%	0.932[f]
Spending (4)	Medicare Part A & B spending ($/beneficiary/month)	Average over intervention quarters 5–9	All observable Medicare FFS beneficiaries	-2.4%	17.6	28.1	$1,150	119.7 (70.6)	11.6%	0.886[e]
			All observable high-risk Medicare FFS beneficiaries	-15.0%	56.3	94.5	$2,499	210.8 (238.1)	9.2%	0.640[e]
		Average over intervention quarter 9	All observable Medicare FFS beneficiaries	-2.6%	13.5	17.7	$1,218	194.9 (150.6)	19.0%	0.781[e]
			All observable high-risk Medicare FFS beneficiaries	-15.0%	25.3	48.0	$2,316	166.0 (524.3)	7.7%	0.507[e]
	Combined (%)			-8.8%	26.4	50.7	n.a.	n.a.	11.9%	0.812[f]

Table II.C.4 *(continued)*

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] Statistical power is the probability of concluding (in this study, using a one-sided test, and a $p < 0.10$ threshold) that the program had a favorable effect, when the true effect was of the specified size. The power calculation is based on actual standard errors from analysis. For example, in the first row, a 5.0 percent effect on all-cause admissions (from the estimated counterfactual of 107.8 - 7.4 = 100.4) would be a change of 5.0 admissions. Given the standard error of 8.4 from the regression model, we would be able to detect a statistically significant result 24.6 percent of the time if the impact was truly 5.0 admissions, assuming a one-sided statistical test at the $p = 0.10$ significance level.

[b] The counterfactual is the outcome the treatment group beneficiaries would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[c] Percentage difference is calculated as the regression-adjusted difference-in-differences estimate, divided by the estimate of the counterfactual.

[d] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test.

[e] We adjusted the *p*-values from the primary test results for the multiple (four) comparisons made within the domain.

[f] This *p*-value tests the null hypothesis that the mean difference-in-differences estimate across the four outcomes in the domain, expressed as percentage change from the estimated counterfactual, is greater than or equal to zero (a one-sided test).

ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

n.a. = not applicable.

d. Results for secondary tests

Overview. Like the primary tests, the secondary tests are conducted using the quarterly data presented in Tables II.C.3.a and II.C.3.b, but with outcomes combined across multiple quarters and regression-adjusted for (1) matched-set-by-quarter fixed effects, (2) differences in beneficiaries' characteristics within each matched set, and (3) the differences that existed between the treatment and comparison groups during the baseline period.

We present the results of our secondary tests in Tables II.C.5.a and II.C.5.b. Overall, the results from the secondary tests suggest the primary test results might reflect differences between the treatment and comparison groups that were *not* caused by the 21st Century Care program.

Results of tests for parallel trends during the baseline period. Our difference-in-differences models assume that the treatment and comparison trends would have been parallel during the intervention period in the absence of the program. This assumption is untestable. However, to determine whether the assumption is plausible, we tested the hypothesis that trends were parallel during the *baseline* period.

Table II.C.5.a shows that trends for the treatment and comparison groups diverged substantively for all outcomes during the baseline period, although the differences generally are not statistically significant. This suggests that the assumption of parallel trends in the absence of the intervention might be—but is not necessarily—violated. For example, for outpatient ED visits, the regression-adjusted treatment group rate increased (worsened) relative to the estimated counterfactual over the course of the baseline period by 5.3 percent among the full population and by 15.4 percent among the high-risk subgroup. The substantive threshold values for the primary tests of ED visits (respectively, 5.0 and 15.0 percent) are not directly comparable (because the numbers for divergence reflect differences in trends during the baseline period rather than projected eventual differences during the primary test period), but still provide a benchmark for judging whether the magnitude of divergence is meaningful from a policy perspective. For all outcomes shown in Table II.C.5.a, the differences in baseline trends between the treatment and comparison groups are meaningfully large. However, the sign (direction) of these differences varies across outcomes. Further, no differences are statistically significant in a two-sided test (that is, testing for either improvement or worsening of the treatment group relative to the estimated counterfactual) when the Type 1 error rate[1] is set to 5 percent. Only one difference (of six) is significant using a Type I error rate of 10 percent, and most are not close to significant. The large *p*-values for most differences, despite their large magnitude, reflects the fact that even relatively large differences could be due to chance.

[1] In statistical tests, a Type I error occurs when we reject the null hypothesis (calling a result statistically significant) when the null hypothesis is true. Both 5 and 10 percent are conventional thresholds for the Type I error rate—that is, thresholds for calling an effect estimate statistically significant.

Table II.C.5.a. Results of secondary tests for Denver Health: Testing for parallel trends in the baseline period

Outcome	Population	Substantive threshold for primary test	Divergence over the six baseline quarters (treatment – estimated counterfactual)[a]	*p*-value[b] (two-sided test)
All-cause inpatient admissions	All Medicare FFS beneficiaries	-5.0%	-15.0%	0.31
Outpatient ED visits	All Medicare FFS beneficiaries	-5.0%	5.3%	0.73
Medicare Part A and B spending	All Medicare FFS beneficiaries	-2.4% or 2.6% (depending on time period)	-4.7%	0.72
All-cause inpatient admissions	High-risk Medicare FFS beneficiaries	-15.0%	-15.4%	0.39
Outpatient ED visits	High-risk Medicare FFS beneficiaries	-15.0%	15.4%	0.42
Medicare Part A and B spending	High-risk Medicare FFS beneficiaries	-15.0%	-28.4%	0.10

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The results for each outcome are based on a regression model estimating the linear trend in the outcome over time during the baseline period, with the linear trend differing by treatment status (that is, with an interaction term between time and an indicator of whether the beneficiary is a member of the treatment group). The covariates and matched-set-by-quarter fixed effects are the same as those used in the difference-in-differences model to estimate impacts, described in the text.

[a] The percentage divergence between the treatment group and the estimated counterfactual (or between the high-risk subgroup of the treatment group and its estimated counterfactual) over the six baseline quarters is calculated as the cumulative difference in treatment and comparison trends, as estimated in the regression model, divided by the treatment group mean in the first baseline quarter (B1).

[b] *p*-values test the null hypothesis that the regression-adjusted difference between the treatment group trend and the estimated counterfactual is zero (a two-sided test).

ED = emergency department, FFS = fee-for-service.

Results of tests for impacts during I1 through I4 (first year of intervention period). We estimated impacts during the first year of the intervention period, when we expected true program impacts to be very small, because large observed impacts during this period could indicate that differences between the treatment and comparison groups are driven by something other than the 21st Century Care program. Table II.C.5.b shows substantively important differences between the treatment group outcomes and the estimated counterfactual for all service use outcomes (specifically, for admissions and ED visits among both the full population and the high-risk subgroup). These differences show that treatment group outcomes were 13.7 to 19.9 percent higher than the estimated counterfactual. In contrast, the estimated impact of 21st Century Care on spending was smaller than the substantive thresholds during the first year of the program. Spending was 1.0 percent higher for the treatment group than for the estimated counterfactual among the full population in I1 through I4 (compared with a substantive threshold in the primary test periods of 2.4 percent in I5 through I9 or 2.6 percent in I9 only; see Table II.C.4). Spending was 3.2 percent higher for the treatment group than the estimated counterfactual in I1 through I4 for the high-risk subgroup, compared with a substantive threshold of 15 percent.

Results of tests restricting analysis to people not participating in the FQHC Demonstration. Table II.C.5.b also shows results of tests for the outcomes and time periods of the primary tests, but restricting to the population that never visited a clinic participating in the FQHC Demonstration. Unadjusted mean outcomes for this restricted population are more variable from quarter to quarter than for the full population (results not shown), most likely because the sample sizes are smaller. However, the sign and rough magnitude of the impact estimates from I5 through I9 or in I9 are not different from those in the primary test results, shown in Table II.C.4.

e. Consistency of quantitative estimates with implementation findings

The primary test results probably cannot plausibly be interpreted as program impacts given the implementation evidence. The magnitude of the primary test estimates is too large and the secondary test results show that estimated impacts appear too soon for these impacts to be attributed to the HCIA-funded program. For example, although it is possible the program could have caused an increase in the outpatient ED visit rate if, say, the newly hired patient navigators advised beneficiaries to go to the ED for unexpected health problems, these patient navigators were not adequately integrated during the first year of the program to have immediate effects. However, our secondary test results show a regression-adjusted difference between the treatment group and the estimated counterfactual of 13.7 percent already during the first year of the program (Table II.C.5.b). Furthermore, this difference was observed among the full population, not only those who would have been receiving services from patient navigators. It is difficult to imagine the program could cause an effect on ED visits—either favorable or unfavorable—among the full population that was as large as the estimated effect of 21.1 percent during I5 through I9 (Table II.C.4).

Table II.C.5.b. Results of secondary tests for Denver Health: Secondary time periods and populations

Secondary test definition				Results			
Domain	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual (standard error)[a]	Percentage difference[b]	*p*-value[c]
Service use	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 1–4	All observable Medicare FFS beneficiaries	112.1	15.8 (7.3)	16.4%	0.985
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			367.6	44.2 (24.4)	13.7%	0.965
	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)		All observable high-risk Medicare FFS beneficiaries	328.6	54.5 (26.5)	19.9%	0.980
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			1,245.8	185.6 (104.0)	17.5%	0.963
	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 5–9	Observable Medicare FFS beneficiaries not affected by the FQHC Demonstration	107.7	4.9 (9.9)	4.8%	0.690
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			442.1	70.3 (40.4)	18.9%	0.959
	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)		Observable high-risk Medicare FFS beneficiaries not affected by the FQHC Demonstration	245.1	26.3 (35.7)	12.0%	0.770
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			1,290.5	205.2 (157.1)	18.9%	0.904
Spending	Medicare Part A & B spending ($/beneficiary/month)	Average over intervention quarters 1–4	All observable Medicare FFS beneficiaries	$1,061	10.6 (61.9)	1.0%	0.847
			All observable high-risk Medicare FFS beneficiaries	$2,555	79.5 (196.2)	3.2%	0.601
		Average over intervention quarters 5–9	Observable Medicare FFS beneficiaries not affected by the FQHC Demonstration	$1,209	128.9 (96.3)	11.9%	0.910
			Observable high-risk Medicare FFS beneficiaries not affected by the FQHC Demonstration	$2,463	394.6 (336.8)	19.1%	0.879
		Intervention quarter 9	Observable Medicare FFS beneficiaries not affected by the FQHC Demonstration	$1,311	242.4 (221.3)	22.7%	0.863
			Observable high-risk Medicare FFS beneficiaries not affected by the FQHC Demonstration	$2,568	690.0 (791.0)	36.7%	0.808

Table II.C.5.b *(continued)*

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] The counterfactual is the outcome the treatment group beneficiaries would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[b] Percentage difference is calculated as the regression-adjusted difference-in-differences estimate, divided by the estimate of the counterfactual.

[c] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test. Values are not adjusted for multiple comparisons.

ED = emergency department; FFS = fee-for-service; FQHC = Federally Qualified Health Center; HCIA = Health Care Innovation Award.

n.a. = not applicable.

f. Conclusions about program impacts, by domain

Based on available evidence, we cannot yet draw conclusions about program impacts (Table II.C.6). The results from the secondary tests and the lack of consistency between the primary test results and implementation evidence both suggest we need more sensitivity checks and, perhaps, refinements to the analytic sample to ensure the model assumptions are met. We plan to draw conclusions about program impacts in future reports.

g. Next steps

All results presented in this report are preliminary. We have several plans to advance the impact evaluation of Denver Health.

First, we will consider refinements to the analytic sample to ensure (1) that model assumptions are met, including that the comparison group is a fair representation of the counterfactual; and (2) that our treatment group comprises beneficiaries with an adequate probability of receiving program services. The following are examples of possible refinements we could make:

- We will consider limiting the analytic sample to beneficiaries observed in all baseline and intervention quarters to assess the extent of selection over time. Because attrition from the treatment and comparison groups is high, it is possible that the groups observed in the intervention quarters differ in important ways from those observed in the baseline quarters. This change in population composition might violate the difference-in-differences assumption that the difference in outcomes for the treatment and comparison groups during the baseline period is stable and would have persisted during the intervention period, were it not for the HCIA-funded program.
- We will consider redefining the high-risk subgroup, recognizing that—according to new research from Denver Health (Johnson et al. 2015)—beneficiaries who were frequent users of acute care services when they first entered our analytic sample might not remain frequent users of acute care services over time. If beneficiaries included in our high-risk subgroup for analysis were, in fact, no longer targeted by Denver Health for intensive services by the time period of our primary tests (starting in I5), we might have limited ability to detect true impacts among high-risk beneficiaries.
- We will examine the sensitivity of the results to outlier observations for all outcomes.

Second, we will examine the possibility that policy changes other than the HCIA and the FQHC Demonstration were occurring at the same time as the HCIA-funded intervention and could have affected service use or spending among Medicare beneficiaries differently in Denver than elsewhere in Colorado. Such policy changes could, conceivably, make our comparison group a poor counterfactual for the Denver Health treatment group, even though the two groups were well matched before entering the analytic sample.

Finally, if data become available, we might add Medicaid FFS beneficiaries to the analytic sample. This would make the analytic sample more representative of the population served by

Denver Health's HCIA-funded program and reduce the degree of imprecision in our impact estimates.

Table II.C.6. Preliminary conclusions about the impacts of Denver Health, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion[a]		
		Primary test results	Primary test result plausible given secondary tests?	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	Not assessed			
Service use	None[a]	The mean estimated effect across all tests in the domain was substantively large and showed an unfavorable difference between the treatment and comparison groups	TBD	TBD
Spending	None[a]	The mean estimated effect across all tests in the domain was substantively large and showed an unfavorable difference between the treatment and comparison groups	TBD	TBD

Source: Tables II.C.4 and II.C.5.

[a] More information is needed to determine whether the primary test results are plausible given the secondary test results and the implementation evidence. We will conduct additional analyses before making conclusions about program impacts in the future.

TBD = to be determined.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

Denver Health received HCIA funding to transform its primary care delivery system to more effectively meet its patients' medical, behavioral, and social needs. Its 21st Century Care program aimed to improve patients' health outcomes, increase patients' satisfaction with between-visit care, and decrease the cost of care. To accomplish these goals, Denver Health developed enhanced primary care teams, established three new high-risk clinics for the system's patients with the most complex conditions, and used health IT to enhance between-visit care. Implementation facilitators included empowerment of frontline staff to adapt implementation strategies based on patients' needs, a commitment to collaboration across multidisciplinary teams, and a system wide emphasis on self-monitoring and continuous quality improvement. Barriers to implementation included difficulties integrating patient navigators into care teams and the complexity of medical needs and resource constraints of the target population. The HCIA-Primary Care Redesign Clinician Survey found that most clinicians believed 21st Century Care would have a positive effect on the quality and patient-centeredness of care and clinicians' ability to respond in a timely way to patients' needs.

Based on the preliminary evidence reported here, we are unable to draw conclusions about program impacts on patients' outcomes in service use or medical spending. The primary test results in both domains showed large unfavorable differences between Medicare FFS beneficiaries assigned to the treatment and comparison groups during the primary test periods (that is, starting either one or two years after the program began). However, the results of secondary (robustness) tests—which found possible divergence of treatment and comparison outcomes during an 18-month baseline (pre-intervention) period, along with large differences during the first program year, before large effects were anticipated—suggest that observed differences might not be caused by the 21st Century Care program. The implementation evidence collected thus far does not provide a clear rationale for why the program would have consistently unfavorable impacts.

Our next steps for this evaluation are to (1) monitor Denver Health's program implementation reports through June 30, 2015, and plans for sustaining the program beyond the funding period; (2) evaluate trainees' and clinicians' attitudes and experiences with the program in the third year of the award through administered surveys; (3) conduct additional robustness checks for the impact evaluation, as described in Section II.C.5.g; (4) extend the impact evaluation to include the full period of program operations and, if possible, include Medicaid FFS beneficiaries; and (5) use the implementation findings to help interpret the impact results.

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Chronic Conditions Data Warehouse. "Table A.1. Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014c. Available at http://www.healthindicators.gov/Indicators/Medicare-beneficiaries-eligible-for-Medicaid-percent_317/Profile/ClassicData. Accessed August 4, 2015.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Johnson, Tracy L, Deborah J. Rinehart, Josh Durfee, Daniel Brewer, Holly Batal, Joshua Blum, Carlos I. Oronce, Paul Melinkovich, and Patricia Gabow. "For Many Patients Who Use Large Amounts of Health Care Services, the Need Is Intense Yet Temporary." *Health Affairs,* vol. 34, no. 8, 2015, pp. 1312–1319. doi:10.1377/hlthaff.2014.1186

Lewin Group. "CMS Health Care Innovation Awards, Round One, Eleventh Quarterly Reporting Period (11QR)." Prepared for the Centers for Medicare & Medicaid Services. June 2015.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21. doi:10.1214/09-STS313.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Finger Lakes Health System Agency

March 2016

Rachel Shapiro
Randall Blair
Rebecca Coughlin
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

FINGER LAKES HEALTH SYSTEMS AGENCY

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by Finger Lakes Health Systems Agency (FLHSA) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the FLHSA program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on patient outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF FLHSA

FLHSA received a three-year, $26.6 million dollar HCIA to implement its program, Transforming Primary Care Delivery: A Community Partnership (hereafter referred to as the FLHSA program or the program). Table I.1 summarizes key features of the program. The goal of this initiative is to transform primary care in 68 practices in six counties in the greater Rochester, New York area. The FLHSA program includes two key components: (1) working with practices to become patient-centered medical homes (PCMHs) and (2) intensive care management of high-risk patients. A third component of the program is to develop a communitywide outcomes-based payment model. Through this program, FLHSA aims to improve intermediate health outcomes and quality of care for high-risk Medicare and Medicaid beneficiaries and, in turn, lower the cost of care by 3 percent. It expects to achieve this goal through a reduction of hospital admissions and readmissions by 25 percent and avoidable emergency department (ED) visits by 15 percent by the end of the award period. FLHSA received a 12-month no-cost extension, during which it will continue to implement all components; its HCIA award will end in June 2016.

Table I.1. Summary of FLHSA PCR program

Awardee's name	Finger Lakes Health Systems Agency
Award amount	$26,584,892
Implementation date	September 2012
Award end date	June 30, 2016
Program description	1. Redesign primary care processes, culture, and workforce to transform practices into PCMHs 2. Train care managers and community health workers to facilitate improved health of high-risk participants and link them with community resources 3. Develop a communitywide outcomes-based payment model
Innovation components	Care coordination, care management, care transitions, medical homes, workflow or process redesign, provider payment reform, patient decision support or shared decision making, population management
Intervention focus	Practice
Workforce development	Hire care managers at each of the participating practices, hire community health workers to work with a subset of practices, redefine responsibilities of existing practice staff
Target population	High-risk Medicare and Medicaid beneficiaries (although all patients are affected)
Program setting	Provider-based (68 practices)
Market area	Regional (six counties in the greater Rochester, New York, area)
Market location	Urban (Rochester), suburban (Webster), and rural

Table I.1 *(continued)*

Core outcomes	• 3 percent reduction in cost of care • 25 percent reduction in hospital admissions and readmissions • 15 percent reduction in avoidable emergency department visits

Source: Review of FLHSA program reports, March 2015.

Notes: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health IT systems, and undertaking other operational activities.

The program was originally scheduled to end on June 30, 2015. However, FLHSA received a no cost extension for 12 months.

The intended target population of this program is high-risk Medicare and Medicaid beneficiaries. However, care managers work with all patients in the practice, regardless of payer and insurance status. To ensure that practice populations reflect the target population, FLHSA used the number of Medicare and adult Medicaid patients as one of its assessment criteria when identifying practices for the program. As of June 2015, the majority of direct participants (70 percent) were Medicare and Medicad beneficiaries (The Lewin Group. Quarterly Awardee Performance Report: FLHSA, 12th Quarterly Reporting Period. Prepared for CMMI, August 2015.).

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how it has been adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environment. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of FLHSA's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in April 2014 and April 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

The FLHSA program has two main structural components (Table II.A.1). First, FLHSA staff work with participating practices to redesign primary care processes, culture, and workforce—for example, by creating care teams, integrating care managers into practice care teams, and implementing care team huddles—to transform 68 practices (recruited in three separate cohorts) into PCMHs. The National Committee for Quality Assurance (NCQA) already recognizes some participating practices as PCMHs; however, the practices still require technical assistance to continue to improve. Other practices are not recognized as PCMHs by NCQA or other organizations. Achieving PCMH recognition is not an explicit goal of this program. However, if practices are interested in obtaining such recognition, participating in program activities might help them do so. . As part of the program's transformation effort, FLHSA practice improvement

advisors help practices to collect and use data to identify areas for practice improvement and test new ideas in practice using a Plan-Do-Study-Act model (Langley et al. 2009). Second, FLHSA clinical advisors help participating practices to train and deploy care managers to provide intensive care management and link patients with community resources. Until December 2014, Trillium Health worked with six practices with a large proportion of high-need patients to integrate community health workers (CHWs) into the practices. CHWs were no longer integrated into practices after this time because of challenges related to integrating them into the practices (see Section II.A.3.b for more detail). The CHWs helped care managers in these practices to link patients with community resources. In addition to these two structural components, over the course of the program, FLHSA leadership has worked with two insurers to develop a communitywide outcomes-based payment model to ensure sustainability of program activities and personnel after the HCIA period.

Table II.A.1. Key details about program design and adaptation

	Program component	
	Practice transformation	**Care management**
Target population	68 practices: • Cohort 1: 19 practices began February 2013 • Cohort 2: 29 practices began July 2013 • Cohort 3: 20 practices began July 2014	High-risk Medicare and Medicaid patients in the 68 participating practices, although care managers work with any patient identified as high risk, including patients who are uninsured.
Patient identification	Not applicable	Care managers screen practice populations to identify high-risk patients who qualify for intensive care management services by using a screening tool (such as the LACE Index Scoring Tool for Risk Assessment of Hospital Readmissions [Van Walraven et al. 2010]); reviewing practice population data; reviewing medical records to find patients with recent hospitalizations or ED visits; receiving a provider's recommendation; and through patient self-referral.
Patient and practice recruitment and enrollment	FLHSA staff: • Conducted outreach to practices in the target area • Assessed practices on four criteria: (1) number of Medicare and adult Medicaid patients, (2) use of EHRs for at least six months, (3) number of primary care providers, and (4) award readiness • Conducted interviews and scored practices to determine which to include	After care managers identify patients as high-risk, they reach out to patients to explain care management and invite them to participate and, if they agree, obtain patients' consent. There is no formal enrollment process.
Service delivery protocol	FLHSA practice improvement advisors work with practice champions and other staff in weekly or biweekly meetings to identify and work on quality improvement projects; team-based care and process improvement concepts are incorporated into each project.	FLHSA clinical advisors: • Coach and mentor practice-based care managers in regularly scheduled meetings to integrate the care manager into the care team at the practice; these meetings initially took place weekly, and now take place less frequently.

Table II.A.1 *(continued)*

	Program component	
	Practice transformation	**Care management**
	All projects rely on the Plan-Do-Study Act model and quality improvement process to test change, document processes, and communicate lessons learned.	• Suggest that care management caseloads should start at 20 to 40 intensive patients in their first year at the practice and gradually build to 40 to 60 patients by the second year. • Recommend that care managers contact patients at least monthly, but do not provide guidance about whether there should be additional contact. Care managers work with patients until they feel patients would no longer benefit from their services or patients decide they no longer need care management.
Adaptations	FLHSA followed a less formal practice recruitment strategy for Cohorts 2 and 3 than for Cohort 1; it focused on practices' willingness to participate in program	In December 2014, because of challenges related to the integration of CHWs into six participating practices, FLHSA stopped working with Trillium Health to identify and employ CHWs in these practices. Over time, some practices targeted a broader set of high-risk patients than was initially identified

Sources: Interviews from second site visit, April 2015; document review, March 2015; Van Walraven et al. 2010.

Note: Primary care payment reform is a supplemental component and is not listed in the table.

EHR = electronic health record.

FLHSA provides participating practices with financial incentives to support transformation activities. Practices receive stipends and funding for care managers, both of which are allocated to practices based on the size of each practice's patient panel and adjusted for risk. For the first three years of the program, practices received support based on their risk-adjusted panel sizes, ranging from a minimum annual stipend of $40,000 and funding for the equivalent of half of a full-time care manager position to a maximum annual stipend of $100,000 and funding for the equivalent of two full-time care managers. During the no-cost extension period, practices will receive $40,000 to $60,000, depending on the number of care managers in the practice. In the first three years, FLHSA provided practices with a lump sum for transformation activities, but during the extension period, FLHSA will reimburse practices for achieving specific transformation milestones.

b. Target populations and patient identification, recruitment, and enrollment

The intended target population of this program is high-risk Medicare and Medicaid patients. However, care managers work with all patients in the practice, regardless of payer and insurance status. To ensure that practice populations reflect the target population, FLHSA used the number of Medicare and adult Medicaid patients as one of its assessment criteria when identifying practices for the program. As detailed in Table II.A.1, care managers screen practice populations to identify high-risk patients in a variety of ways, such as a screening tool, medical record review, provider recommendation, or patient self-referral. Over time, some practices altered their

patient identification strategies, and went from serving a narrow set of patients (for example, those with diabetes) to a wider set of high-risk patients (such as those with a recent hospital discharge or congestive heart failure).

As described in Table II.A.1, FLHSA staff conducted outreach to recruit practices and assessed their readiness for participation before enrolling them into the program. The participating practices vary in structure and affiliation—they are private practices, Federally Qualified Health Centers (FQHCs), or part of a larger health system. They also vary in terms of the characteristics of their patient populations, such as race and ethnicity, age, comorbid conditions, and coverage source.

c. Intervention staff and workforce development

FLHSA houses program staff who administer day-to-day program implementation activities, contribute to the development of program strategy, and support program self-monitoring. To oversee program implementation, FLHSA hired an HCIA-funded program director, a data analyst, and a program assistant (all of whom are in full-time positions). These staff members work with existing staff at FLHSA to administer specific program activities. In addition, as described in Table II.A.2, practice improvement advisors and a practice improvement coordinator work with practice staff on practice transformation activities, and clinical advisors and a social worker/resource coordinator work with care managers to provide them with guidance and help integrate them into practice care teams.

At the practice level, a number of staff facilitate program implementation. A primary care provider at each practice serves as a practice champion, working as a liaison between FLHSA and practice staff and serving as an advocate for practice transformation and integrated care management. Each practice hired care managers; a few practices have more than one care manager, and many practices have part-time or shared care managers. Most care managers are registered nurses or licensed practical nurses; a few are social workers. In addition, before December 2014, CHWs worked at six practices with a large proportion of high-need patients.

FLHSA staff provide a variety of training and workforce development activities at participating practices. When first hired, care managers attended a comprehensive training on fundamental skills, such as data collection and entry; they receive supplementary trainings on specific topics or skills, as needed. Care managers and practice champions also attend monthly learning collaboratives, which provide opportunities to share lessons and challenges and to learn from the experiences of their peers (either other care managers or practice champions). In October 2014, FLHSA staff also convened two separate joint collaboratives for care managers and practice champions—one for Cohort 1 and another for Cohort 2. FLHSA staff convened a Cohort 3 joint collaborative in March 2015. Attendance at all learning collaboratives is mandatory and practices are penalized financially for repeated poor attendance.

Table II.A.2. Key details about intervention staff

Program component	Staff member	Staff /team responsibilities	Adaptations
Practice transformation	FLHSA practice improvement advisors	• Provide technical support to practice champions • Assess individual practice needs and work with practice staff to develop and test solutions (for example, assisting practices with Plan-Do-Study-Act cycles) • Work with practices to help them transform into PCMHs (for example, identifying processes and resources for managing admissions, discharges, and transitions of patients)	None
	FLHSA practice improvement coordinator	• Oversees practice improvement advisors • Serves as practice improvement advisor for designated practices	None
	Practice champions	• Oversee on-site implementation of practice transformation activities • Serve as main point of contact with FLHSA program staff • Meet regularly with FLHSA practice improvement advisor • Attend FLHSA learning collaboratives	None
Care management	FLHSA clinical advisors	• Provide technical support to care managers (for example, helping care managers report on clinical quality measures through practice EHRs) • Meet with each care manager at least biweekly to discuss challenges and provide education and training on topics such as motivational interviewing, EHR use, and care team relationships (for example, building rapport with other staff at the practice)	• Clinical advisors were not in the initial staffing plan; FLHSA added these positions after the program began. • As an alternative to individual meetings with care managers, clinical advisors are piloting small group meetings (3 to 8 care managers grouped by practice affiliation with a health system or medical group).
	FLHSA social worker/resource coordinator	• Provides resources and technical assistance to care managers to help connect patients with necessary services at community-based service organizations • Organizes trainings and networking sessions to introduce care managers to community resources	None

Table II.A.2 *(continued)*

Program component	Staff member	Staff /team responsibilities	Adaptations
	Care managers	• Provide intensive care management to high-risk patients • Work with practice staff to define the embedded care management role and implement care management processes, such as daily huddles and weekly care team meetings • Communicate with practice providers by documenting care they provide in EHRs, discussing the patients at care team meetings and huddles, and meeting informally with the providers during the workday • On a monthly basis, care managers submit data to FLHSA about their patients and the care services delivered, such as number of patients on their caseload, PAM scores, and insurance information • Meet regularly with FLHSA clinical advisor and social worker • Attend FLHSA learning collaboratives	Vision of care manager's role evolved: • Initially, FLHSA clinical advisors expected that care managers would spend 35 percent of their time on intensive care management, 25 percent on population management, 30 percent on care transitions, and 10 percent on developing relationships in the practice. • Over the course of the program, the clinical advisors revised this guidance after recognizing that the care manager's role encompasses more than these areas and will continue to evolve as the practice transforms more completely into a PCMH.
	Community health workers	• Educate practice staff in six practices on the needs of the local community • Connect patients to external resources	• In December 2014, because of challenges related to the integration of CHWs into the six practices, FLHSA stopped working with Trillium Health to identify and employ CHWs in these practices.
All components	Program director	• Oversees program strategy and execution, manages program staff and relationships with external partners, conducts research, and disseminates findings	None
	Data analyst	• Analyzes clinical and financial data, and obtains, collects, and analyzes data for program use	None
	Program assistant	• Provides administrative and logistical support to program and program staff	None

Sources: Interviews from second site visit, April 2015; document review, March 2015.

Note: Primary care payment reform is a supplemental component, and is not listed in the table.

EHR = electronic health record; PAM = Patient Activation Measure.

In addition to the learning collaboratives for practice staff, FLHSA staff also convene trainings for practice improvement advisors and clinical advisors. Until early 2015, FLHSA staff held weekly coaching sessions to improve team development and communication skills; these sessions ceased when staff felt the sessions became less effective. In mid-2015, FLHSA staff were collaborating with a physician leadership consultant to develop a coaching workshop—with

a training curriculum and toolkit—to improve their ability to coach practices through leadership development and practice transformation.

d. Service delivery protocols

As detailed in Table II.A.1, FLHSA practice improvement advisors follow a project-based model when working with practices. Many practices first focused on establishing communication pathways among practice staff and, specifically, among care teams in the practice. Then, practices focused on projects such as improving workflows, process mapping, cycle time analysis, and defining a health coach position.

FLHSA clinical advisors are coaches and mentors for care managers, helping them to implement care management processes at the practice level and integrate the care managers into practice care teams (Table II.A.1). Initially, clinical advisors worked with care managers to establish care teams and regular huddles, if they were not already being held, as well as regular care team meetings. At the initial and subsequent meetings, clinical advisors discuss with care managers how to identify patients for care management and build and maintain an intensive care panel, as well as population management. Additional content of the meetings varies based on the needs of each care manager and practice. Over the course of the program, FLHSA clinical advisors also have helped care managers to focus on serving high-risk patients with complex care needs (for example, those with chronic obstructive pulmonary disease, congestive heart failure, or diabetes).

Care management services vary depending on patient needs, but follow a similar structure in all practices. After patients are identified as high–risk and in need of intensive care management (described in Table II.A.1 and Section II.A.1.b), care managers reach out to them to explain care management and invite them to participate. If patients agree to participate, care managers meet with them—by telephone, in person at the practice during regular medical appointments or specific care management appointments, and occasionally through home visits—to help manage their diseases and prevent hospital readmissions, conduct medication reconciliation, and coordinate care and referrals to social services in the community. Throughout these interactions, which take place at least monthly, care managers use motivational interviewing, education, and teach-back opportunities to help engage patients in their own care. Care managers use a Patient Activation Measure (PAM; developed by Insignia Health) to assess patients' activation to improve their health; the patient's activation is assessed three times—at the first care management visit, 90 days after beginning care management, and at discharge. Care managers use the PAM scores to help them assess patient needs and their continued need for intensive care management. In addition to using the PAM scores, interviewed care managers reported that they assess patient needs on a case-by-case basis, saying that they might contact patients requiring more guidance, such as those recently discharged from the hospital, as often as once a day, depending on their needs. As patients become more capable of caring for themselves, care managers start to move them off of their panels and stop providing them with care management services.

2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators and self-reported information included in FLHSA's quarterly self-monitoring and measurement reports.

a. Program enrollment

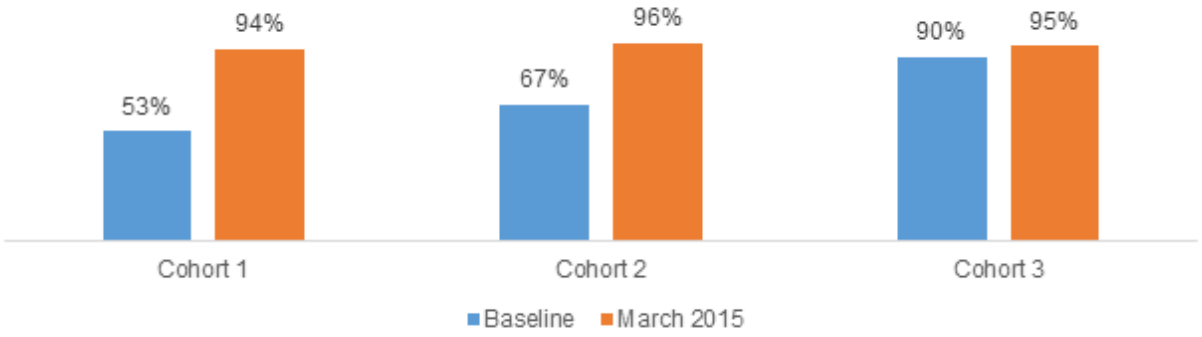
FLHSA successfully enrolled practices and reached more patients than expected. FLHSA recruited 68 practices (exceeding its target of 65). As of December 2014, perhaps because of the larger number of practices participating, FLHSA care managers provided services (both intensive and otherwise) to 14,472 unique patients, exceeding the target cumulative enrollment of 13,564 care managed patients (about half of whom were expected to receive intensive care management) for the entire award period.

b. Service measures

FLHSA's self-monitoring metrics indicate that practices are effectively transforming the way they deliver care, and have improved on a variety of process measures in the short time (9 to 27 months, depending on the cohort) they have participated in the program. Next, we describe FLHSA's self-reported metrics for each of the three cohorts separately, as each cohort has a different baseline (defined as the first month the practices were enrolled)—January 2013 for Cohort 1, July 2013 for Cohort 2, and July 2014 for Cohort 3. Although Cohort 1 practices have had the most time to make progress, FLHSA staff noted that Cohorts 2 and 3 benefited from lessons learned early in the program.

By March 2015, practices reported that they were using their electronic health records (EHRs) more effectively to manage care. Although FLHSA did not establish a target for EHR use among participating practices, its practice improvement advisors work with practices to improve their use over time. As of March 2015, 95 to 100 percent of practices in all three cohorts were using their EHRs to generate population-based reports sorted by patient ages and major diagnoses, increasing from baseline measures of 47 percent (Cohort 1), 80 percent (Cohort 2), and 95 percent (Cohort 3). In addition, as of March 2015, participating practices increased their use of EHRs to generate and use patient-specific reports to identify gaps in care (Figure II.A.1).

Figure II.A.1. FLHSA self-reported percentages of practices using EHRs to generate patient-specific reports, by cohort



Source: FLHSA. Eleventh Quarter Measuring and Monitoring Results. Prepared for CMMI, March 2015.

Note: This information is based on the awardee's self-reported data. We have not attempted to verify its completeness or quality. The first month of program participation (baseline) varies by cohort. Month 1 is January 2013 for Cohort 1, July 2013 for Cohort 2, and July 2014 for Cohort 3.

Practices also started to transform their practices into PCMHs and develop operational care teams by March 2015. FLHSA practice improvement advisors and clinical advisors worked with practices to improve practice team communication through monthly team meetings and weekly huddles. Figure II.A.2 illustrates the practices' incorporation of monthly care team meetings and weekly huddles over the course of the award. As of March 2015, according to FLHSA, more than 90 percent of practices in the first two cohorts reported holding at least one care team meeting per month; almost two-thirds (65 percent) of Cohort 3 practices reported holding monthly care team meetings. All practices reported holding at least one huddle or planning session each week.

Figure II.A.2. FLHSA self-reported percentage of practices holding monthly care team meetings and weekly huddles, by cohort and month of program participation

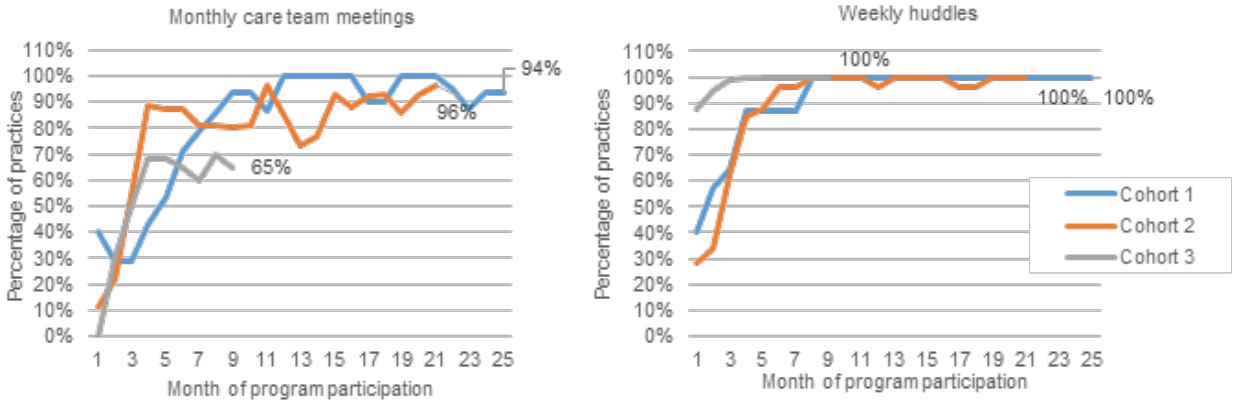


Figure II.A.2 *(continued)*

Source: FLHSA. Eleventh Quarter Measuring and Monitoring Results. Prepared for CMMI, March 2015.

Note: This information is based on the awardee's self-reported data. We have not attempted to verify its completeness or quality. The first month of program participation varies by cohort. Month 1 is January 2013 for Cohort 1, July 2013 for Cohort 2, and July 2014 for Cohort 3. Labeled percentages indicate the percentage of practices in each cohort holding monthly care team meetings or weekly huddles as of March 2015.

c. Staffing measures

FLHSA successfully met its staffing goals for the program. Although FLHSA experienced some staff turnover, as of April 2015, it employed 11 program staff. At FLHSA, these staff included 5 practice improvement advisors and 1 practice improvement coordinator (for 6 total, an increase from the original 3), 4 clinical advisors (a role added in response to practice needs), and 1 social work clinical coordinator. Participating practices also experienced some turnover in care managers, but as of April 2015, each practice met FLHSA's goal to employ at least 1 care manager, resulting in a total of 70 embedded care managers across the 68 practices. FLHSA hired 6 CHWs through its partner, Trillium Health, but as of December 2014 all 6 CHWs had left because of challenges related to integrating them into the practices (discussed further in Section II.A.3.b).

FLHSA reported that many practice staff participated in workforce development activities—the initial care manager training and the monthly learning collaboratives. Most attended either the initial trainings in February 2013 (Cohort 1), July and August 2013 (Cohort 2), and July 2014 (Cohort 3), or the make-up trainings held from August 2013 to December 2014. Cohort 1 care managers attended five consecutive, day-long training sessions (40 hours total). Because these care managers reported feeling overwhelmed by the amount of training delivered in five consecutive days, FLHSA staff decided to break up the Cohort 2 care manager training into a pair of two-day sessions (32 hours total). Care managers hired later in the process attended two 8-hour make-up sessions. As of December 2014, FLHSA reported that it had trained a total of 107 care managers (including care managers who left the participating practices and those who replaced them). In addition, as of March 2015, FLHSA reported that more than 80 percent of care managers and practice champions in all three cohorts attended their respective learning collaboratives. These high rates are not unexpected, as participation in the collaboratives is a condition of receiving funding from FLHSA.

d. Program time line

FLHSA successfully implemented two of the planned program components on time for all three cohorts of practices—practice transformation and intensive care management. FLHSA faced challenges in implementing its third component—developing payment models offered by two commercial insurers, Excellus Blue Cross Blue Shield and MVP Health Care—by the original target of July 2014. As of April 2015, MVP had developed a payment model and Excellus was in the process of finalizing its model. As described further in Sections IV.A.3.d and IV.A.4, the burgeoning accountable care organization (ACO) market hindered the development of these payment models.

In May 2015, FLHSA received a 12-month no-cost extension. During this extension, FLHSA staff will continue to support practices from all three cohorts financially and with technical assistance.

3. Implementation experience

In this section we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.3 summarizes the major facilitators of and barriers to FLHSA's implementation effectiveness in each domain.

Table II.A.3. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Perceived relative advantage • Adaptability	• No significant barriers noted
Implementation process	• Self-monitoring/quality improvement • Staff engagement	• Execution • Program resources
Internal factors	• Leadership characteristics • Team characteristics • Implementation climate	• No significant barriers noted
External environment	• Payment models: synergy with ACO practice transformation activities	• Payment models: streamlining program with ACO care management requirements

Source: Interviews from second site visit, April 2015; document review, March 2015.

a. Program characteristics

Two characteristics stood out as facilitators of program implementation: (1) the perceived relative advantage of the program compared with the standard delivery of care and (2) the adaptability of the program to practices' and patient needs. First, in interviews, practice staff reported that several factors improved their care delivery since they began participating in the program, including: an increased emphasis on the care team, the presence of an embedded care manager, and improved communication with patients. As a result of the program's focus on team-based care, practice staff reported that they had either begun holding or increased the frequency of huddles, had improved the efficiency and effectiveness of previsit planning, and were adapting to a team-based approach to care. In particular, practice staff appreciated the collaboration provided through the care team approach, viewing it as an advantage over the way they previously provided care. As one practice champion noted during our site visit in April 2015, "Involving the nurse, care manager, myself … we're much more involved about communicating about each patient, and hopefully having the patient be involved in the team." The care teams are facilitated by the care managers, who also supplement care delivered by providers; they accomplish this by meeting with patients to clarify providers' instructions and connect patients with community resources. In addition, they communicate with providers about any identified patient needs. Interviewed providers commented that they appreciated the added

degree of patient-centered care delivered by the care managers, which they felt had led some patients to better control their conditions.

Second, practice staff's ability to adapt the program to their own practices' needs helped them to implement the program effectively. This adaptability is built into the FLHSA program design; FLHSA practice improvement advisors and clinical advisors tailor their coaching and mentoring to the needs of the practices and care managers. In transforming practice workflows, FLHSA practice improvement advisors allowed practices to chart their own course, identifying projects that would help them achieve more patient-centered care. These projects generally started with implementing daily huddles and moved on to such topics as previsit planning, creating new office protocols, and using EHR data for quality improvement. FLHSA also allowed practices to use the approaches that worked best to identify and reach their targeted high-risk patients and provide them with care management. As shown in Table II.A.1, FLHSA allowed practices to use different strategies to identify high-risk patients; some practices also relied on their health system or ACO to help identify patients who needed intensive care management. This often resulted in practices providing care management to different populations: for example, some of the visited practices targeted patients with uncontrolled diabetes, whereas others targeted a broader group of patients with a range of chronic conditions. In addition, FLHSA practice improvement advisors and clinical advisors did not limit practices to implementing a standardized model of care management; instead, they allowed practices to assess their patients' needs and tailor their use of the care manager in a way that best met their patients' needs. This resulted in care managers providing services in a variety of ways; although most visits occurred in person at the practice and over the telephone, some interviewed care managers also reported providing care in patients' homes.

b. Implementation process

Two implementation process factors facilitated FLHSA program implementation: (1) using data to self-monitor and conduct quality improvement activities and (2) staff engagement related to the embedded care manager role. FLHSA helps practices to monitor their own progress, as well as how they compare with other participating practices' progress, by providing them with quarterly reports summarizing practice-level clinical, quality, and cost data; these quarterly reports supplement any reports that practices generate through their EHRs or receive from their hospital system or ACO. Partly as a result of these activities, interviewed providers reported seeing more staff engagement with the embedded care manager. Indeed, those we interviewed reported that staff engagement increased over the course of practices' participation in the program. At first, respondents reported that some staff hesitated to embrace care managers, largely because they did not understand how the care managers should function in the practices. As providers grew to understand the care managers' role, saw them in action, and noticed changes in some patients' behaviors, providers started to appreciate the added care being provided and were more likely to refer high-risk patients to the care managers. As of April 2015, all interviewed care managers reported feeling accepted and integrated into the practices. One care manager commented during our April 2015 site visit, "When I first got here it was a lot about education, teaching them about what I do and how I can help them. Some of doctors were a little more reserved. They've really opened up as time's gone on."

Two implementation process factors presented challenges to program implementation: (1) program execution in relation to integrating CHWs into practices and (2) program resources in relation to the time required for transformation activities. First, although Trillium Health and FLHSA staff initially worked with selected practices to integrate CHWs, in hindsight it is clear that communications could have been improved, as several of these practices were unclear on how CHWs should function in their practices and did not assign work to the CHWs. One practice—whose staff felt they were already connecting patients with community services—decided it did not need the additional help of a CHW and eliminated the position. FLHSA staff felt these issues could have been prevented by more effective management of the relationship between the practices and the CHWs. In addition, FLHSA staff felt Trillium and FLHSA should have engaged practices earlier in the process to ensure staff better understood how the CHWs should function in their offices—for example, as extensions of the care manager and conducting home visits to assess such issues as fall risks and medication compliance. During our April 2015 site visit, one practice improvement advisor commented, "The lack of an individual needs assessment from each individual practice from the start, and the lack of engagement in the planning and development of what it was led to this unfortunate situation. Trillium, the vendor, had a clear idea of what they thought CHWs did from their experience, but [that] didn't necessarily match everyone else's." FLHSA leadership suggested that the integration of CHWs into practices might have been improved in part by providing focused trainings and mentoring for CHWs and practices to clarify the CHW role and expectations of the position.

Second, practice champions and other providers struggled to devote sufficient time to the transformation activities, reporting that it was difficult to find time in their busy schedules to attend daily huddles, care team meetings, and learning collaboratives. Care managers also reported struggling to find sufficient time to perform all of the tasks required because of their high caseloads and the requirements that came with participating in the FLHSA program (such as entering data about the patients receiving care management and attending the learning collaboratives). The interviewed care managers reported that they often worked overtime to document the care they provided—both for the practice (in EHRs) and FLHSA. To address this issue, FLHSA staff reported that they worked with care managers on how to appropriately manage their time and caseloads.

c. Internal factors

Characteristics of the organization implementing a program can influence implementation effectiveness. Three internal factors facilitated implementation of the FLHSA program: (1) leadership commitment, (2) team characteristics, and (3) implementation climate. First, practice champions and practice managers were committed to transforming their practice workflows; improving communication among members of the care team; and integrating care managers into the care team, particularly in light of national and statewide initiatives for new payment models based on the provision of patient-centered care and quality improvement. At the visited practices, staff pointed to the practice champions as a driving force behind practice change.

Second, practices demonstrated strong team communication and collaboration. Much of FLHSA's coaching related to practice transformation focuses on building successful care teams. As a result, and perhaps not surprisingly, interviewed practice staff reported that these care teams

helped move the practices forward in their transformation efforts. During our April 2015 site visit, one practice champion said, "I've learned that you should definitely have a team approach. Bringing all disciplines together to understand that specific patient is very helpful … what we can do to better enhance patient care, by not being complacent, to have set goals, have [the] team be able to bring forward ideas and suggestions is really important. That collaboration between everybody is really important."

Third, interviewed practice staff reported implementation climates that were favorable to practice transformation and integration of care managers. Most staff were committed to improving how they provided care and worked collaboratively with care managers. Some practices were already moving toward becoming a PCMH before participating in the FLHSA program; at these practices, staff readily embraced team huddles and the opportunity to practice at the top of their licenses. Although staff at other practices were initially apprehensive about integrating a care manager into their teams, according to interviewed FLHSA staff, many became more accepting of the position as they learned how care managers could ease the burden on providers and lead to improved health outcomes among the practices' patients.

d. External environment

Features of an organization's external environment can also influence program implementation. External payment models developed by two regional ACOs, Accountable Health Partners (AHP) and the Greater Rochester Independent Practice Organization (GRIPA), facilitated the implementation of the FLHSA initiative. Since the beginning of the FLHSA program, many participating practices joined one of these ACOs. Interviewed practice staff commented that the ACOs' support of practice transformation—by providing practices with population data or consultants to assist with PCMH certification, as well as providing some financial support for care management—helps to spur them forward in their practice transformation efforts. According to one practice champion interviewed during our April 2015 site visit, the ACO had "made a commitment there, so now it's a daily routine. Culture has changed."

At the same time, external payment models also presented challenges; ACOs' system requirements for care manager caseloads and working with high-risk patients are sometimes stricter than FLHSA's requirements. FLHSA practice improvement advisors and clinical advisors reported that they were working closely with the ACOs to streamline guidance for care management such that FLHSA guidance did not conflict with system requirements. Their end goal was to ease the burden on practice staff, who reported that they were making their own decisions about how to follow the two sets of directives from FLHSA and the ACOs.

4. Sustainability and scalability

The ability to sustain the HCIA practice transformation in participating practices will depend on (1) practices' commitment to the changes they have made and (2) whether practices will be reimbursed for the services care managers provide—either through a payment model developed by the insurers or through their ACOs. According to FLHSA staff we interviewed during our April 2015 site visit, practices fully support practice transformation and most "don't

ever want to go back. They like this model, they like the team-based care, they see the value of the care manager, they see the value of proactively managing the panel." Practice champions and other respondents expected they would sustain huddles and previsit planning because they now view these activities as integral to the practice workflows. Practices and FLHSA staff recognized, however, that staff commitment to the transformation changes would not necessarily be enough to sustain the changes. They felt that the key to sustaining practice transformation and dedicated care management would be reimbursement. If they do not receive financial support to employ a care manager, several interviewed practice staff felt that care management services at the practice would dwindle—largely because providers and their support staff would find it too time-consuming to more directly manage patient care. As FLHSA staff noted, "The finance thing is the 200 pound gorilla in the room. If it's not financially sustainable, it's not sustainable. These after all are businesses. There isn't some well they can dip in and just take money out of. They have to be able to pay their staff and make payroll."

Financial support for embedded care managers after the program ends also depends on the practice's ownership status and ACO membership. FLHSA leadership initially planned to work with two commercial insurers—Excellus BlueCross BlueShield and MVP Health Care—to develop a communitywide payment model to ensure the sustainability of the practice transformation and care management activities begun under the FLHSA program. However, because many practices joined one of two regional ACOs, they no longer rely on the payment model being developed by Excellus and MVP. Of the two ACOs in the Rochester region, one (AHP) will pay for the salaries and benefits of all embedded care managers at its member practices for 6 to 12 months after the end of the program, at which point the ACO will reevaluate whether to continue to provide this support. The second ACO (GRIPA) will support a centralized, telephonic care management model. Embedded care managers in GRIPA's system-owned practices will receive support from the ACO for an unspecified period of time; however, private practices face losing their dedicated care managers and will instead have to rely on remote (telephonic and non-office-based) care managers. Those practices that are not part of an ACO expect to receive shared savings for performance on specified quality and outcome measures through the Excellus and MVP payment models. MVP's model is developed and is in the pilot stage, and Excellus is working with FLHSA leadership to finalize its model, which Excellus expects to be in place by the end of the program. FLHSA leadership expect that the combined shared savings payments to practices will cover continuing practice transformation costs and the cost of employing a care manager.

Participating practices might also receive additional financial support by billing for care management costs. Over the past year, FLHSA staff researched how Medicare fee-for-service codes that allowed practices to bill for chronic care management (which went into effect in January 2015) could be used and whether they would benefit the participating practices. If, over the course of the next year, practices express interest in using these codes, FLHSA practice improvement advisors will provide them with technical assistance. However, both FLHSA and practice staff expressed reservations about the viability of these codes. Among the principle concerns were that using the codes would require patients to agree to a copayment, and the documentation requirements could potentially be time-consuming enough to cancel out the revenue received by billing for the codes.

As of April 2015, program activities were already being spread to patients and practices not participating in the program—largely among practices that were part of either a larger group of practices or a health system. Among the visited practices, staff reported that their larger medical groups were investing in broader PCMH initiatives and building on the transformation activities already taking place as part of the FLHSA program. Some practice champions reported that, based on the perceived success of these practice transformation activities, other (nonparticipating) practices that were part of the same medical group were adopting or would soon adopt huddles and care team meetings. According to some practice staff, health systems also expressed an interest in adopting care management services like those in the participating practices; some systems engaged care managers from participating practices to train new care managers at practices that are not participating in the FLHSA program.

After the end of the HCIA, FLHSA staff hope to be able to provide practices that participated in the FLHSA program, as well as other practices in the region, with continued guidance on practice transformation activities. FLHSA leadership expects to apply for external funding so that FLHSA can remain a resource center for practices in the region.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected clinical sites or satellite offices provided important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff were limited to a small number of practices and might not reflect the perspectives of clinicians practicing at other sites. In order to assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in fall 2014, the third year of the HCIA-funded program. Data from the survey provide additional insights into the implementation process and experience as well as the contextual factors that might affect implementation effectiveness in the FLHSA program.

In this section, we report on clinicians' views of their daily work life and practice. First, we focus on the contextual factors that might affect program implementation, including the characteristics of the practice locations, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care team functions. We then present data on the alignment of the clinicians' views and experiences with the overall goals of the HCIA-funded innovation, as well as their awareness of and participation in the FLHSA program and their view of the facilitators of and barriers to successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

A total of 86 clinicians at practices participating in the FLHSA program responded to the survey (resulting in a response rate of 70 percent). Of the 86 respondents, 61 were physicians, 11 were nurse practitioners, and 13 were physician assistants. (The number of clinicians in each

response category [here and throughout this section] does not always sum to the total number of respondents because of survey item nonresponse or questions that might not apply to all respondents.) These clinicians practiced predominantly at clinical locations with three or more clinicians (64 percent). Other clinical practice sites included FQHCs (17 percent) or some other practice type (16 percent). Clinicians reported that their primary source of compensation was a salary adjusted for performance (50 percent), a fixed salary (16 percent), fee for service (14 percent), or another source (14 percent).

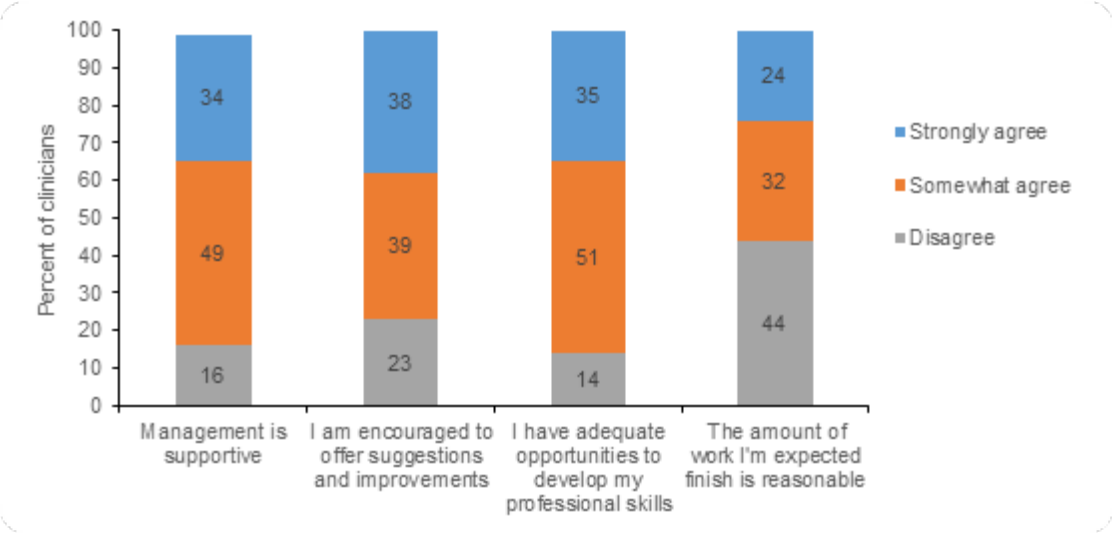
Clinicians at participating FLHSA practices reported working in settings that are advanced in terms of health IT. Although nationally slightly more than half of physicians practice in settings with functional EHRs (Furukawa et al. 2014), most clinicians reported using health IT at their practice locations. Most clinicians reported using electronic systems for drug dosing and interaction alerts (95 percent), prescribing medications (94 percent), entering clinical notes (94 percent), accessing laboratory test results (92 percent), or ordering tests and procedures (84 percent). In addition, about 80 percent of clinicians reported using electronic referral tracking systems and patient registries, functions that are generally advanced and not in widespread use nationally (DesRoches, Painter, and Jha 2014). More than half of clinicians also reported that they offer patient-facing technologies, providing their patients with the option of doing the following online: request a prescription refill (67 percent), request an appointment (57 percent), and email a clinician about a medical question or concern (57 percent).

Most clinicians reported that their practices are focused on improving quality of care. In addition to the FLHSA program, 73 percent of responding clinicians reported participating in a collaborative quality improvement effort with other organizations in the past two years. Almost two-thirds of all respondents (64 percent) said they received training on quality improvement and tools in the past two years, and 74 percent said they conducted at least one clinical audit of the care their patients receive.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall careers, levels of burnout, and perceptions of their practice environment can all have an effect on the success of program implementation and organizational change. The majority (85 percent) of responding clinicians reported being generally satisfied with their careers in medicine. However, about one-third (35 percent) were experiencing some symptoms of burnout at the time the survey was taken. Clinicians at practices participating in the FLHSA program gave consistent ratings to their workplace management. As shown in Figure II.B.1, more than one-third of responding clinicians strongly agreed that their management team was supportive, that they were encouraged to offer suggestions and improvement, and that they had adequate opportunities for professional development. About one-quarter strongly agreed that the amount of work they were expected to complete each day was reasonable, which is consistent with the site visit finding that time was a barrier to program implementation.

Figure II.B.1. Workplace ratings



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Clinicians generally reported spending their time doing work that was appropriate to their level of training, but reported spending some time doing things they felt others with less training could do. Slightly fewer than half of respondents (48 percent) reported that they spent at least three-quarters of their time doing work that is well matched to their training. A majority said that they spent less than one-quarter of their time doing work for which they did not have enough training (89 percent) or that the work could be done by someone with less training (55 percent).

In addition to workplace ratings, the survey included items that assess clinicians' beliefs about their ability to provide high quality care. Two-thirds of responding clinicians either strongly or somewhat agreed with the statement "It is possible to provide high quality care to all of my patients." The majority of clinicians reported that major barriers to providing optimal care were lack of time to spend with patients, insufficient reimbursement, lack of timely information about care provided to patients by other physicians, patients' inability to pay for care, difficulties obtaining specialized diagnostic tests or treatments, and difficulties obtaining specialist referrals.

c. Clinicians' perceptions of care team functioning

The vast majority of clinicians (91 percent) in practices participating in the FLHSA program reported working as part of a care team and, overall, their perceptions of how these teams function were positive. Most clinicians agreed that members of the care team relayed information in a timely manner (94 percent), had sufficient time for patients to ask questions (90 percent), used common terminology when communicating with one another (88 percent), verbally verified information they received from one another (66 percent), and followed a standardized method of sharing information when handing off patients (61 percent).

d. Alignment with goals of PCR

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The views of clinicians in practices participating in the FLHSA program generally aligned with the goals of PCR (Table II.B.1). A majority of clinicians rated 7 of the 13 goals as extremely important. However, only 26 percent of FLHSA program clinicians rated "increasing the number of primary care practices functioning as a patient-centered medical home" as extremely important and only 39 percent rated "improving the capability of health care organizations to provide team-based care" as extremely important. This is notable because the FLHSA program's practice transformation efforts are driven by supporting practices to function as PCMHs with a focus on team-based care.

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the FLHSA program is to change the way care is provided, with one focus being to redesign the primary care practice workforce. Clinicians are critical to that process, and understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on patient outcomes. For example, if clinicians are aware of the program, have received appropriate and effective training, and believe that the FLHSA program will have a positive effect on the care they provide, they are likely to feel more invested in the program's success. Alternatively, those who feel more negatively about the program might be less likely to implement the intervention enthusiastically. In this section, we report on clinicians' experiences with and perceptions of the FLHSA program.

Table II.B.1. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Increasing access to primary care	59	67%
Improving care continuity in primary care	56	64%
Improving appropriateness of care	55	63%
Improving patients' capacity to manage their own care	53	60%
Improving care coordination for patients with chronic conditions	52	59%
Reducing ED visits	51	58%
Reducing hospital readmissions	47	53%
Increasing the use of evidence-based practices in clinical care	43	49%
Reducing overall health care spending	42	48%
Improving capability of health care organizations to provide team-based care	34	39%
Improving the capability of health care organizations to provide patient-centered care	30	34%
Increasing the number of primary care practices functioning as a PCMH	23	26%
Increasing the use of EHRs and other health IT	23	26%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Notes: Percentages are calculated as the percentage of total respondents who rated each item as "extremely important." Items are rated separately; percentages do not add up to 100 percent.

a. Awareness of the program and receipt of training

Among the clinicians we surveyed, 82 percent were at least somewhat familiar with the FLHSA program. Of these clinicians, 71 percent had received training related to the program. On average, clinicians received 13.5 hours of program-related training. Almost two-thirds (64 percent) of surveyed clinicians reported that they had been at their practice for more than five years; 32 percent had been at the practice between one and five years, and 2 percent had been at the practice for less than one year.

b. Perceived effect of program on patient care

Clinicians' perceptions of the effect of the FLHSA program on the care they provide to patients were largely positive. (Clinicians were asked about the perceived effect of the FLHSA program and the barriers to and facilitators of implementation only if they reported being at least somewhat familiar with the program.) As shown in Table II.B.2, most clinicians who were familiar with the FLHSA program believed it would have a positive effect on the patient-centeredness and quality of the care they provide, their ability to respond to patient needs in a timely way, and safety. Fewer than half of physicians familiar with the FLHSA program believed it would have a positive effect on equity or efficiency.

Table II.B.2. Perceptions of effects of program on patient care

	Positive effect		Negative effect, no effect, or too soon to tell	
Survey item	Number	Percentage of respondents	Number	Percentage of respondents
Patient-centeredness	52	70%	21	29%
Quality of care	48	65%	26	35%
Ability to respond in a timely way to patient needs	44	59%	30	41%
Safety	40	54%	34	46%
Equity of care for all patients	33	45%	41	55%
Efficiency	28	38%	46	63%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Figures are based on the total number of FLHSA program clinicians reporting they were at least somewhat familiar with the FLHSA program.

c. Barriers to and facilitators of program implementation

Finally, we asked clinicians who were at least somewhat familiar with the FLHSA program to rate the effect of a series of barriers and facilitators to program implementation. The level of program funding, the quality of interpersonal communication with other allied health professionals, and the availability of community resources to care for patients with complex conditions were the most often-cited facilitators to program implementation. In addition, more than half of the surveyed respondents saw the availability of personnel, availability of relevant patient information at the point of care, and quality of interpersonal communications with other providers as having a positive effect on implementation. Clinicians also perceived the required

use of computer and communications technology and the availability of evidence-based clinical information as having a positive effect.

4. Conclusions about clinicians' attitudes and behavior

Clinicians at practices participating in the FLHSA program were generally positive about their jobs and ability to provide high-quality primary care. Despite some reports of burnout and room for improvement in the amount of time spent doing work that could be done by someone with less training, clinicians reported feeling supported by management and supportive of the FLHSA program and its goals. In particular, clinicians felt that the program has the potential to improve the quality of care and patient-centeredness of care. High percentages of clinicians named program funding, quality of interpersonal communications with other allied health professionals, and availability of community resources as facilitators of program implementation. These are consistent with the FLHSA program's efforts to integrate primary care with community services and support. Additionally, a majority of clinicians agreed that all elements of care teams were functioning, suggesting that FLHSA's emphasis on team-based care is permeating practices and providing further evidence that the FLHSA program is starting to transform care delivery. Findings from the HCIA Primary Care Redesign Clinician Survey support information learned during the site visit, and suggest that the opinions shared by interviewed providers represent the group of clinicians at participating practices more broadly.

C. Impacts on patient outcomes

1. Introduction

In this part of the report, we draw preliminary conclusions, based on available evidence, about the impacts of FLHSA's HCIA program on patients' outcomes in three domains: quality-of-care outcomes, service use, and spending. We first describe the methods for estimating impacts (Section II.C.2) and then the characteristics of HCIA program practices (treatment practices) at the start of the intervention (Section II.C.3). We next demonstrate that treatment practices were similar at the start of the intervention to the practices we selected as a comparison group, which is essential for limiting potential bias in impact estimates (Section II.C.4). Finally, in Section II.C.5, we describe the quantitative impact estimates, their plausibility given implementation findings, and our conclusions about program impacts in each domain.

Our conclusions in this report are preliminary because they are based on outcomes for Medicare beneficiaries assigned to practices in the first two of three cohorts participating in the HCIA program, and because these analyses do not yet cover the full time over which the intervention is expected to have an effect. In future reports, we plan to include Medicaid beneficiaries assigned to these practices, to extend the outcome period to cover the full length of the intervention, and to include additional practices that began the intervention as part of the third cohort of program practices. Finally, although the FLHSA program serves a mix of Medicare fee-for-service (FFS) beneficiaries, Medicare beneficiaries enrolled in managed care plans, Medicaid beneficiaries, and patients with other forms of insurance, due to limitations in available data we have analyzed outcomes only for the Medicare FFS population (including those who are dually eligible for Medicare and Medicaid). Results might not be generalizable to the full population that FLHSA program serves.

2. Methods

a. Overview

We estimated program impacts as the difference in outcomes between Medicare beneficiaries assigned to treatment practices and matched comparison practices, adjusting for any pre-intervention differences between the groups. We estimated impacts for two distinct populations: the full Medicare patient population and the high-risk Medicare population served by program practices. We included both of these populations in the analysis because FLHSA anticipated reducing utilization and costs among its full patient population as a result of the HCIA program, but expected to generate most of these reductions by improving care for beneficiaries at high risk of hospitalization and other expensive care.

We estimated program impacts using a difference-in-differences framework. To implement this framework, we first calculated the average difference in outcomes for Medicare beneficiaries assigned to treatment practices before they joined the intervention (the baseline period) and after they joined (the intervention period). Then, we subtracted from this difference the average difference in outcomes for Medicare beneficiaries assigned to matched comparison practices during the baseline versus intervention periods. This approach helps to isolate program impacts from any differences in outcomes for the treatment and comparison practices before the start of the intervention or temporal changes that equally affected both groups over the study period.

Our impact estimates capture the combined effect of the program's two key components: (1) working with practices to become PCMHs and (2) intensive care management of high-risk patients. The estimates do not include the impacts of the program's third component—development of a new payment model that supports practice transformation—because that payment model was introduced in 2015, after the analysis period covered in this report. In addition, these estimates do not include the impacts of the program on non-Medicare beneficiaries who could potentially benefit from the intervention.

b. Treatment group definition

The treatment group includes 38 of the 48 practices that joined the HCIA program in the first two cohorts: 16 cohort 1 practices and 22 cohort 2 practices. We excluded 10 practices for two reasons. First, for 6 practices, we were unable to find suitable comparison practices. This included 5 FQHCs and one practice providing primary care to patients admitted to the hospital for psychiatric conditions. Second, for the remaining 4 practices, we attributed fewer than five Medicare FFS beneficiaries to the practice during the baseline period, making it impossible to implement the difference-in-differences design for these practices. We also excluded cohort 3 practices from this analysis because neither we nor the awardee expected effects for these practices during the outcome period covered by this report. We anticipate including cohort 3 practices in future reports that have longer follow-up periods. We could also include participating FQHCs in future analyses if, by expanding the geographic region from which comparisons can be drawn, we find suitable comparison practices.

Because the practices joined the intervention in cohorts, the time period for the baseline and intervention period varies by cohort. For the 16 cohort 1 practices, the baseline period is January

to December 2012, and the intervention period is January 2013 to December 2014, the end of the outcome period for this report. For the 22 cohort 2 practices, the baseline period is June 2012 to May 2013, and the intervention period is from June 2013 to December 2014. As a result, we were able to report outcomes for four baseline quarters for all 38 practices (across both cohorts), eight intervention quarters for cohort 1 practices, and six intervention quarters for cohort 2 practices.

To focus our analyses, we specified a limited number of primary tests before examining any impact results. Each primary test defined an outcome, population, time period, and direction of expected effects for which we hypothesize to see impacts if the program is effective. We provided the awardee and CMMI an opportunity to comment on the primary tests, and revised them as appropriate. We drew conclusions about impacts in each domain based on the results of these primary tests and the consistency of the primary test results with the implementation findings and secondary quantitative tests (which test the robustness of assumptions underlying the primary test results).

We defined the treatment group separately in each quarter in the baseline and intervention periods. For each quarter in each period, the treatment group consists of Medicare FFS beneficiaries who (1) were assigned (see below) to one of the treatment practices on or before the first day of the quarter, (2) had observable outcomes for at least one day in the quarter, and (3) lived in New York or Pennsylvania for at least part of the quarter. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, have Medicare as their primary payer, and are not enrolled in a comprehensive managed care plan.

Beneficiary assignment to practices. The first step involved attributing beneficiaries to practices using the same decision rule that CMMI uses for the Comprehensive Primary Care (CPC) initiative. Specifically, in each baseline and intervention month, we attributed beneficiaries to the primary care practice whose providers (physicians, nurse practitioners, or physician assistants) provided the plurality of primary care services in the past 24 months. When there was a tie, we attributed the beneficiary to the practice he or she visited most recently. This attribution method requires identifiers for the providers who worked in the treatment practices (and when) as well as identifiers for providers in other practices in the region who could compete for patients (when determining which practice provided the plurality of primary care services). FLHSA provided identifiers for the treatment providers and SK&A, an outside health care data vendor, supplied identifiers for providers in the other practices. Second, in each period (baseline and intervention), we *assigned* each beneficiary to the first treatment practice he or she was attributed to in the period, and continued to assign him or her to that practice for all quarters in the period.

Using this definition of the treatment group, a beneficiary who has previously been assigned to the treatment group would *remain* a member of the treatment group for the rest of the relevant period (baseline or intervention), as long as he or she was still enrolled in Medicare FFS and living in New York or Pennsylvania by the end of the relevant period. This definition ensures that during the intervention period, beneficiaries do not exit the treatment group solely because the intervention was successful in reducing their service use (including visits at program

practices). The definition for the baseline period then corresponds to that of the intervention period so that, across the two periods, interpretation of the population changes over time should be comparable.

Definition of a high-risk subgroup. Because some aspects of FLHSA's intervention (including care management) focus on improving care for beneficiaries at high-risk of hospitalization and other expensive care, we also defined a high-risk subgroup of the treatment group each quarter. For each baseline quarter, this subgroup consists of the beneficiaries with a Hierarchical Condition Category (HCC) score in the top quarter among all observable treatment group members at the start of the baseline period. The HCC score, developed by the Centers for Medicare & Medicaid Services (CMS), is a continuous variable that predicts a beneficiary's Medicare spending in the following year relative to the national average, with 1.0 indicating that the predicted spending is at the national average and 2.0 indicating that it is twice that average. In each intervention quarter, the high-risk subgroup consists of beneficiaries whose HCC scores were in the top quarter among all observable treatment group members at the start of the intervention period.

c. Comparison group definition

The comparison group consists of Medicare FFS beneficiaries assigned to 77 matched comparison practices—54 cohort 1 comparison practices and 23 cohort 2 comparison practices—during each quarter of the baseline and intervention periods. We selected comparison practices that were similar during the baseline periods to the treatment practices in factors that can influence patients' outcomes, especially those that FLHSA used when deciding which practices to recruit for the intervention. This section describes how we constructed the matched comparison group whereas Section II.C.4 shows the balance we achieved between the two groups on the matching variables.

We selected the 77 comparison practices in four steps:

First, we limited the potential comparison practices to the approximately 2,000 primary care practices in New York State that were located (1) outside of the greater New York City area and (2) outside of the six counties in which FLHSA is operating. This formed the initial population of primary care practices that could feasibly be matched to treatment practices based on practice and patient characteristics. We excluded New York City because the demographics and market characteristics in New York City are very different from the rest of the state. We excluded primary care practices in the six treatment counties because FLHSA has recruited many of the practices in those counties, and the remaining practices that are not participating could systematically differ from those that are (for example, in interest in participating in practice transformation activities).

Second, we constructed matching variables, defined prior to the start of the intervention for all treatment and potential comparison practices. These variables include characteristics of the practices (for example, the number of primary care providers [PCPs] in the practice and the practice's EHR use) as well as characteristics of all Medicare FFS beneficiaries assigned to the practices (for example, mean HCC score, Medicare Part A and B spending, and utilization in the

baseline period); and characteristics of high-risk beneficiaries assigned to the practices. (Section II.C.4 provides additional detail on matching data and results.) We developed a cohort 1 and a cohort 2 version of matching variables for each potential comparison practice—with different one-year baseline periods for each version—so that they could be matched to either a cohort 1 or a cohort 2 treatment practice. As noted earlier, the baseline period was January 2012 to December 2013 for cohort 1 and June 2012 to May 2013 for cohort 2.

Third, we narrowed the pool of potential comparison practices by excluding those practices that: (1) were located in counties in which at least one of two federal primary care initiatives were operating: The Multi-Payer Advanced Primary Care Practice (MAPCP) Demonstration and the CPC initiative; and (2) had an average of fewer than 50 assigned Medicare FFS beneficiaries during the four baseline quarters. These exclusions made the comparison pool better resemble treatment practices, because none of the treatment practices were located in areas in which CPC or MAPCP was operating, and all treatment practices had more than 50 assigned Medicare beneficiaries during the baseline period. These restrictions left a pool of 567 potential comparison practices.

Fourth, we used propensity-score methods to select 77 comparison practices from the pool of 567 that were similar to the 38 treatment practices on the matching variables. The propensity score for a given practice is the predicted probability, based on all matching variables, that the practice is part of the treatment group (Stuart 2010). The score collapses information from all of the matching variables into a single number for each practice that we used to assess how similar practices are to one another. We matched each treatment practice to one or more comparison practice with a similar propensity score, with the aim of generating a comparison group that was similar, on average, to the treatment group on the matching variables (see Section II.C.4 to assess balance between treatment and comparison groups after matching).

We ran two separate propensity-score matching models—one matching cohort 1 treatment practices to comparison practices with characteristics defined over the cohort 1 baseline period, and the other matching cohort 2 treatment practices to comparison practices with characteristics defined over the cohort 2 baseline period. We required each treatment practice to match to at least one, but no more than five, comparison practices. In some cases, we also allowed two treatment practices in the same cohort to match to the same comparison practice. However, to prevent complications in the statistical models, we did not allow a comparison practice to match to a cohort 1 treatment practice *and* a cohort 2 treatment practice. These matching rules resulted in a ratio of comparison to treatment practices that was approximately 2:1. This matching ratio increases the statistical certainty in the impact estimates (relative to 1:1 matching), because it creates a more stable comparison group against which to compare the treatment group's experiences. The matching ratio was higher (3:1) for the cohort 1 practices than for the cohort 2 practices (1:1) because it was harder to find suitable matches for cohort 2 practices.

After selecting the comparison practices, we assigned Medicare FFS beneficiaries to them in each intervention quarter using the same rules we used for the intervention group and for matching (see Section II.C.2.b). Further, we defined a high-risk subgroup of comparison

members in each quarter using the same rules as for the treatment group (that is, using the distribution of HCC scores observed among treatment group members).

d. Construction of outcomes and covariates

We used Medicare claims from January 1, 2009, to December 31, 2014, for beneficiaries assigned to the treatment and comparison practices to develop two types of variables: (1) **outcomes,** defined for each person in each baseline or intervention quarter that the person is a member of the treatment or comparison group; and (2) **covariates** that describe a beneficiary's characteristics at the start of the baseline and intervention periods, and are used in the regression models for estimating impacts to adjust for a beneficiary's characteristics before the period began. We used covariates defined at the start of each period, without updating them each quarter, to avoid controlling in each intervention quarter for previous quarters' program effects, as this would bias the effect estimates away from detecting true impacts. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. We calculated five quarter-specific outcomes that we grouped into four domains:

1. Domain: Quality-of-care outcomes
    a. Inpatient admissions (number/quarter) for ambulatory care-sensitive conditions (ACSCs);
    b. Number of inpatient admissions followed by an unplanned readmission within 30 days (number/quarter)
2. Domain: Service use
    a. All-cause inpatient admissions (number/quarter)
    b. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
    a. Total Medicare Part A and B spending ($/month)

Four of these outcomes—all but admissions for ACSCs—are outcomes that the Center for Medicare & Medicaid Innovation (CMMI) has specified as core for the evaluations of all HCIA programs. Our definition of the readmission measure, however, differs from CMMI's standard definition. CMMI typically defines readmissions as the proportion of inpatient admissions that end in an unplanned readmission. Instead, we analyze impacts on the *number* of these unplanned readmissions per 1,000 beneficiaries per quarter because this enables us to look at the total impact on readmissions across the treatment group, rather than readmissions contingent on an inpatient admission (because the intervention might affect the number and type of admissions as well).

Covariates. The covariates include (1) whether a beneficiary has each of 18 chronic conditions (heart failure, chronic obstructive pulmonary disease, chronic kidney disease,

diabetes, Alzheimer's and related dementia, depression, ischemic heart disease, cancer, asthma, hypertension, atrial fibrillation, stroke, hyperlipidemia, hip fracture, osteoporosis, rheumatoid arthritis, bipolar disorder, and schizophrenia); (2) whether a beneficiary was dually eligible for Medicare and Medicaid; (3) HCC score; (4) demographics (age, gender, and race or ethnicity); and (5) original reason for Medicare entitlement (old age, disability, or end-stage renal disease).

e. Regression model

We used a regression model to implement the difference-in-differences design for estimating impacts. For each quarter-specific outcome, the model estimates the relationship between the outcome and a series of predictor variables, assuming that each of those variables has a linear (additive) relationship with the outcome. The predictor variables include the beneficiary-level covariates (defined in Section II.C.2.d); whether the beneficiary is assigned to a treatment or comparison practice; an indicator for each practice (which accounts for differences between practices in their patients' outcomes at baseline); indicators for each post-intervention quarter; and an interaction of a beneficiary's treatment status with each post-intervention quarter. The estimated relationship between the interaction term and outcomes in a given quarter is the impact estimate for that quarter. It measures the average difference between outcomes for beneficiaries assigned to the treatment and comparison practices that quarter, subtracting out any differences between these groups during the four baseline quarters.

By providing separate impact estimates for each intervention quarter, the model enables the program's impacts to change the longer the practices are enrolled in the program (which is expected to occur). We can also test impacts over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the impact estimates, allowing for statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. The model used robust standard errors to account for clustering of outcomes across quarters for the same beneficiary and a dummy variable for each practice (fixed effects) to account for clustering of outcomes for beneficiaries assigned to the same practice. Appendix 2 provides details on the regression methods, including descriptions of the weights each beneficiary receives in the model.

f. Primary tests

Table II.C.1 shows the primary tests for FLHSA, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important (expressed as a percentage change from the counterfactual—that is, the outcomes the treatment group would have had in the absence of the HCIA-funded intervention). The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for detail and a description of how we selected each test). We provided both the awardee and CMMI an opportunity to comment on the primary tests.

Table II.C.1. Specification of the primary tests for FLHSA

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population	Substantive threshold (impact as percentage of the counterfactual)[d]
Quality-of-care outcomes (4)	Inpatient admissions for ACSCs (#/person/quarter)	Medicare and Medicaid: Average over I5 through I14 (cohort 1); I5 through I12 (cohort 2); and I5 through I8 (cohort 3)[e]	Medicare FFS, Medicaid FFS, and Medicaid managed care beneficiaries assigned to treatment practices	-5.0%
	30-day unplanned readmission rate (#/person/quarter)			-5.0%
	Inpatient admissions for ACSCs (#/person/quarter)		High-risk Medicare FFS, Medicaid FFS, and Medicaid managed care beneficiaries assigned to treatment practices	-15.0%
	30-day unplanned readmission rate (#/person /quarter)			-15.0%
Service use (4)	All-cause inpatient admissions (#/person /quarter)[c]		Medicare FFS, Medicaid FFS, and Medicaid managed care beneficiaries assigned to treatment practices	-2.7%
	Outpatient ED visit rate (#/person/quarter)			-5.0%
	All-cause inpatient admissions (#/person/quarter)[c]		High-risk Medicare FFS, Medicaid FFS, and Medicaid managed care beneficiaries assigned to treatment practices	-4.9%
	Outpatient ED visit rate (#/person/quarter)			-15.0%
Spending (2)	Medicare Part A and B and Medicaid FFS spending ($/person/month)[c]		Medicare FFS and Medicaid FFS beneficiaries assigned to treatment practices	-1.6%
	Medicare Part A and B and Medicaid FFS spending ($/person/month)[c]		High-risk Medicare FFS and Medicaid FFS beneficiaries assigned to treatment practices	-3.3%

Notes: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.
High-risk beneficiaries are defined as those with HCC scores in the top 25 percent among all beneficiaries assigned to the treatment practices (see text for details).

[a] We adjusted the *p*-values from the primary test results for the multiple comparisons made within each domain, but not across domains.

[b] The regression models control for differences between the treatment and comparison groups during the baseline year when estimating program impacts.

[c] For all-cause hospital admissions and Medicare FFS spending (for all patients and high-risk patients), we set the substantive threshold to 75 percent of our estimate of FLHSA's expected effect in each relevant time period (see text for details). For the other outcomes, we set the threshold equal to reductions in acute care use or spending that Peikes et al. (2011) indicated could be feasible among general and high-risk population beneficiaries in a PCMH program.

[d] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention.

[e] To specify the primary tests, we made assumptions about the Medicaid data that will be available by our final report. We believe it is possible we could have Medicaid FFS and managed care data for New York covering the period through June 2016 (I14 for cohort 1, I12 for cohort 2, and I8 for cohort 3). However, this is not guaranteed.

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** FLHSA's central goal is to reduce ED visits, 30-day unplanned readmissions, ACSC admissions, and total medical spending. FLHSA did not explicitly state that it expected to reduced all-cause hospital admissions. However, through the expected reductions in ACSC admissions, FLHSA should also reduce all-cause admissions (although as a smaller percentage change). We plan to assess program effects on all five of these outcomes.
- **Time period.** FLHSA obtained a one-year extension past its original HCIA funding end date of June 30, 2015. Under this extension, FLHSA will continue to implement all components of its HCIA program until the new end date of June 30, 2016. To maximize our ability to detect program impacts, we plan to analyze program impacts from early 2014 to mid-2016 among patients in cohorts 1, 2, and 3. This corresponds to intervention quarters 5 through 14 (I5 through I14) for cohort 1, I5 through I12 for cohort 2, and I5 through I8 for cohort 3. FLHSA officials expect the program to have no effects in a practice's first year of participation, half of the maximum effect in the second year, and the full effects in the third year and beyond. By mid-2016, the first cohort will have experienced 3.5 years of the intervention (with potential for the maximum effect during the last six intervention quarters), the second cohort will have experienced 3.0 years of the intervention (with potential for the maximum effect during the last four intervention quarters), and the third cohort will have experienced 2.0 years of the intervention (with potential for half of the maximum effect during the last four intervention quarters).
- **Population.** FLHSA's impacts should occur among all Medicare and Medicaid patients attributed to the treatment practices, because the intervention is practice-wide. However, impacts should be concentrated among high-risk beneficiaries, both because there are more opportunities to reduce acute care for this high-risk population and because beneficiaries in this group are more likely to receive intensive care management services. Therefore, we plan to include both populations (all beneficiaries and high-risk beneficiaries) in our primary tests. We will assess all outcomes presented above for both the entire Medicare FFS and the high-risk populations.
- **Direction (sign) of the impact estimate.** The primary tests use one-sided tests for *reductions,* relative to the counterfactual, for each of the outcome measures.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we have specified thresholds for what we call substantive importance. We express the threshold as a percentage change from the counterfactual—that is, the outcomes that beneficiaries in the treatment group would have had if they had not received the treatment. For the full patient population, the 2.7 and 1.6 percent thresholds we chose for all-cause hospitalizations and total spending, respectively, are 75 percent of FLHSA's expected effects among all three cohorts during the primary test period (I5 through I14). (We use 75 percent recognizing that FLHSA could still be considered successful if it approached, but did not achieve, its fully anticipated effects.) The 5 percent threshold for the remaining outcomes is extrapolated from the literature (Peikes et al. 2011), which suggests

that impacts of this size should be considered substantial, even though they are smaller than the impacts FLHSA anticipates. (By the third year of the intervention, the awardee expects a decrease of 25 percent in potentially preventable hospitalizations and 30-day hospital readmissions, and a decrease of 15 percent in ED visits among its full patient population.)

For the high-risk patient population, the 4.9 and 3.3 percent thresholds we chose for all-cause hospitalizations and total spending, respectively, are 75 percent of our estimate of FLHSA's expected effects among high-risk beneficiaries for all three cohorts during the primary test period (I5 through I14). This estimate is based on the percentage of high-risk beneficiaries in the population and the portion of utilization and costs for which they account vis-à-vis patients who are not at high risk. The 15 percent threshold for the remaining outcomes is extrapolated from the literature (Peikes et al. 2011), for the same reason as above (that is, the literature indicates effects of this size should be considered substantial, even though they are smaller than our calculation of FLHSA's expected effects for high-risk beneficiaries).

Due to limitations in data availability, we were able to conduct the primary tests in this report only partially. Specifically, we estimated impacts only through I8 for cohort 1 practices and through I5 for cohort 2 practices. As noted earlier, in future reports we plan to cover the full period described in the primary tests (I5 through I14), to include the cohort 3 practices and Medicaid beneficiaries. However, this will require that reasonably current Medicaid data will be available for the final analyses, which is not guaranteed.

g. Secondary tests

We also conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the non-experimental design or random fluctuations in the data. We will have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. Specifically, we estimated the program's impacts on all five outcomes for the full population and high-risk beneficiaries during the first 12 months after the practices joined the intervention (I1 through I4 for both cohorts). Because we and FLHSA expect program impacts to increase over time, with few or no impacts in the first year of practice participation in the program, the following pattern would be highly consistent with an effective program—little to no measured effects in the first four quarters, growing effects in I5 through I8, and the largest impacts in I9 through I14. In contrast, if we found very large differences in outcomes (favorable or unfavorable) in the first 12 intervention months, this could suggest a limitation in the comparison group, not true program impacts.

h. Synthesizing evidence to draw conclusions

Within each domain, we drew one of four conclusions about program effectiveness, based on the primary test results, the results of secondary tests, and the plausibility of those findings given the implementation evidence. These four possible conclusions are as follows: (1) statistically significant favorable effect (the highest level of evidence), (2) substantively important favorable effect, (3) indeterminate effect, and (4) substantively important unfavorable

effect. (We cannot conclude that a program has a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program harms.)

Our decision rules for each of the four possible conclusions are described in Appendix 3. In short, we concluded that a program has a statistically significant favorable effect in a domain if (1) at least one primary test result in the domain was favorable and statistically significant, after adjusting the statistical tests to account for multiple tests (if applicable) within a domain; or (2) the average impact estimate across all primary tests in the domain was favorable and statistically significant. In both cases, we also had to determine that the primary test results were plausible given the secondary tests and implementation evidence. We concluded that a program has a substantively important favorable effect if the average impact estimate was substantively important but not statistically significant, and if the result was plausible given the secondary tests and implementation evidence. In contrast, if the average impact estimate was unfavorable (opposite the hypothesized direction), larger than the substantive threshold, and unfavorable effects were plausible given the other evidence, we concluded that the program had a substantively important unfavorable effect. Finally, if the tests in a domain did not meet any of these criteria, we concluded that the impact in that domain was indeterminate.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the treatment group at the start of the intervention (January 1, 2013, for cohort 1 practices and June 1, 2013, for cohort 2 practices), which can be seen in the second column of Table II.C.2. (Table II.C.2 also serves a second purpose—to show the equivalence of the treatment and comparison practices at the start of the intervention—which we describe in Section II.C.4.)

Characteristics of the practices overall. Our analysis includes 38 treatment practices at the start of the intervention, none of which are FQHCs. Almost all treatment practices had providers receiving payment from CMS for Meaningful Use of EHRs (92 percent). This latter proportion is consistent with FLHSA's targeting, as one of the program's eligibility criteria was an EHR system that practice staff used actively for at least a year. Treatment practices had 6.4 total providers, on average. The large majority of practices' clinicians in the treatment group had a primary care specialty.

Characteristics of the practices' Medicare FFS beneficiaries. The demographic characteristics of all Medicare FFS beneficiaries assigned to the treatment group during the baseline period were, overall, comparable to nationwide FFS averages. Patients in the treatment group also had hospital and ACSC admission rates, 30-day readmission rates, and HCC scores that were comparable to national averages. However, the mean outpatient ED visit rate (135/1,000 people/quarter) was higher than the national average of 105. In part, this might reflect the proportion of dually eligible beneficiaries in treatment practices, which, at 31 percent, is higher than the national average of 22 percent among FFS beneficiaries. People who are dually enrolled in Medicare and Medicaid tend to have higher ED rates than Medicare beneficiaries who are not dually enrolled (Medicare Payment Advisory Commission 2013).

Characteristics of the practices' high-risk Medicare FFS beneficiaries. The high-risk beneficiaries in the treatment group had substantially greater health care needs during the baseline period than the full treatment group (Table II.C.2). Their mean HCC risk score was more than twice the mean for all treatment group members (2.3 versus 1.1), consistent with how the group was defined. Further, they had more than twice the number of all-cause inpatient admissions and Medicare spending than the full population of attributed beneficiaries.

4. Equivalence of the treatment and comparison groups at the start of the intervention

Demonstrating that the treatment and comparison groups are similar at the start of the intervention is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying difference-in-differences models—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment practices not received the intervention.

Table II.C.2 shows that the 38 treatment practices (16 in cohort 1 and 22 in cohort 2) and the 77 selected comparison practices (54 in cohort 1 and 23 in cohort 2) were similar at the start of the intervention on most matching variables. There were some slight differences between the treatment and matched comparison group beneficiaries on the variables we matched through propensity scores, but the standardized differences across the propensity-score matching variables are all within our target of 0.25 standardized differences, and most were within 0.15 standardized differences (the 0.25 target is an industry standard; for example, see Institute of Education Sciences 2014). The omnibus test that the treatment and comparison practices are perfectly matched on all variables cannot be rejected ($p = 0.953$), further supporting that the treatment and comparison groups were similar at the start of the intervention.

We also separately assessed balance among the cohort 1 practices (16 treatment and 54 comparison practices), because these practices change from a subgroup of practices to the full set of practices in later quarters. Specifically, because only cohort 1 practices can be followed up for I7 and I8 for this report, those are the only practices in the treatment and comparison groups in those two intervention quarters. It is important to show that the cohort 1 treatment and comparison practices are balanced at baseline so that regression-adjusted differences in I7 and I8 can be interpreted as program impacts. Among cohort 1 practices, there were some differences between the treatment and matched comparison group beneficiaries on the variables we matched through propensity scores, but the standardized differences across the propensity-score matching variables are all within our target of 0.25 standardized differences, and approximately half are within 0.15 standardized differences (data not shown).

Table II.C.2. Characteristics of treatment and comparison practices when the intervention began

Characteristic	Treatment practices (N = 38)	Unmatched comparison pool (N = 567)	Matched compar-ison group (N = 77)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Exact match variable[c]						
Non-FQHC	100.0	100.0	100.0	0	0	n.a.
Propensity matched variables[d]						
Characteristics of a practices' location(s)						
Located in an urban zip code (%)	82.9	80.6	79.8	3.1	0.08	NA
Zip code poverty rate (%)[e]	14.3	14.7	13.7	0.6	0.05	NA
Located in a health professionals shortage area (primary care)	1.1	1.3	1.1	0.0	0.09	NA
Characteristics of all patients attributed to practices during the baseline year						
Number of beneficiaries	360	401	357	3	0.01	n.a.
HCC risk score	1.12	1.16	1.10	0.02	0.12	1.0
All-cause inpatient admissions (#/1,000 patients/quarter)	76.3	81.6	75.8	0.5	0.02	74[f]
Outpatient ED visit rate (#/1,000 patients/quarter)	135.3	126.4	129.3	6.0	0.10	105[g]
Medicare Part A and B spending ($/patient/month)	697.4	755.6	682.6	14.7	0.10	860[h]
30-day unplanned hospital readmission (#/1,000 patients/quarter)[i]	11.6	11.7	10.3	1.3	0.21	NA
Inpatient admissions for ACSCs (#/1,000 patients/quarter)	14.0	15.6	15.4	-1.4	0.21	11.8[j]
Dually eligible beneficiaries (%)	30.7	20.3	29.5	1.1	0.06	21.7[j]
Disability as original reason for Medicare entitlement (%)	42.7	30.2	41.0	1.7	0.09	16.7[k]
Age (years)	67.2	70.9	67.8	-0.6	-0.10	71[l]
Female (%)	59.8	58.1	58.7	1.1	0.14	55.3[k]
Race: white	82.5	87.6	87.1	-4.6	-0.25	81.8[k]
Characteristics of high-risk patients attributed to practices during the baseline year						
Number of high-risk beneficiaries	86.9	98.3	80.5	6.3	0.08	n.a.
HCC risk score	2.29	2.33	2.27	0.02	0.11	n.a.
Percentage of high-risk beneficiaries	24.2	25.3	23.6	0.6	0.09	n.a.
All-cause inpatient admissions (#/1,000 patients/quarter)	169.4	179.0	166.1	3.3	0.07	74[f]
Outpatient ED visit rate (#/1,000 patients/quarter)	225.1	206.5	205.5	19.6	0.16	105[g]
Medicare Part A and B spending ($/patient/month)	1,481.5	1,535.7	1,443.7	37.7	0.10	860[h]
30-day unplanned hospital readmission (#/1,000 patients/quarter)	32.0	31.8	27.6	4.4	0.22	NA

Table II.C.2 *(continued)*

Characteristic	Treatment practices (N = 38)	Unmatched comparison pool (N = 567)	Matched compar-ison group (N = 77)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Inpatient admissions for ACSCs (#/person/quarter)	37.0	41.0	40.9	-3.9	0.20	11.8[i]
Characteristics of the practices						
Meaningful Use of EHR (%)[m]	92.1	54.7	90.6	1.5	0.04	n.a.
Owned by hospital or health system (%)	53.9	30.4	53.3	0.7	0.01	n.a.
Number of clinicians at practice	6.4	3.9	5.5	0.9	0.12	n.a.
Practices' clinicians with a primary care specialty (%)	93.0	89.7	91.5	1.5	0.07	n.a.
Omnibus test for balance on matching variables[n]						
p-value			0.953			n.a.

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code household income data merged from the American Community Survey ZIP Code Characteristics. Characteristics of the practices come from SK&A, a health care data vendor, as well as the National Committee for Quality Assurance.

Notes: The characteristics for the treatment and their matched comparison practices are defined at the time the treatment practice joined the intervention (January 1, 2013, for cohort 1 practices and June 1, 2013, for cohort 2 practices).

The comparison group means are weighted based on the number of matched comparison practices per treatment practice. For example, if four comparison practices are matched to one treatment practice, each of the four comparison practices has a matching weight of 0.25.

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the matched treatment and comparison groups.

[b] The standardized difference is the difference in means between the matched treatment and comparison groups divided by the standard deviation of the variable, which is pooled across the matched treatment and selected comparison groups.

[c] Exact match means that we required that non-FQHCs match only to non-FQHCs.

[d] Variables that we matched on through a propensity score, which captures the relationship between a practice's characteristics and its likelihood of being in the treatment group.

[e] Average poverty rate associated with each practice's zip code, merged from the American Community Survey.

[f] Health Indicators Warehouse (2014b).

[g] Gerhardt et al. (2014).

[h] Boards of Trustees (2013).

[i] This rate is for individuals ages 65 and older (Truven Health Analytics 2015).

[j] Health Indicators Warehouse (2014c).

[k] Chronic Conditions Data Warehouse (2014a, Table A.1).

[l] Health Indicators Warehouse (2014a).

[m] Meaningful Use of EHRs is calculated as the percentage of practices with at least one provider (National Provider Identifier) working in the practice who received financial incentives for Meaningful Use of certified EHRs through Medicare or Medicaid during the baseline period.

[n] Results from an overall chi-square test indicate the likelihood of observing differences in the matching variables as large as the differences we observed if, in fact, the treatment and comparison populations (from which we drew the samples) were perfectly balanced. The value of $p = 0.953$ for the chi-square test suggests that the two groups are well balanced, because we cannot reject the null hypothesis that their characteristics are identical.

*/**/*** Significantly different from zero at the .10/.05/.01 levels, two-tailed test, respectively. No differences were significantly different from zero at the 0.01 level.

NA = not available.

n.a. = not applicable

5. Intervention impacts

In this section, we first present sample sizes and mean outcomes, by quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates that follow; however, the differences in mean outcomes are not regression-adjusted and not impact estimates by themselves. Next, we present the results of the primary tests, by domain. Then, we present the secondary tests results and assess whether the primary test results are plausible given the secondary tests and given the implementation evidence. We end with preliminary conclusions about program impacts in each domain.

a. Sample sizes

In the first baseline quarter (B1), the treatment group included 13,337 beneficiaries assigned to 38 practices and the comparison group included 31,180 beneficiaries assigned to 77 practices (Table II.C.3). The sample sizes stayed relatively steady across the baseline and the first six intervention quarters, reflecting near balance of two opposing forces—beneficiaries being added to the sample because they are newly assigned to the practices, and beneficiaries dropping from the sample because they died, moved out of the region, or switched from FFS to managed care. As expected, the sum of the comparison group members' weights was roughly equal to the size of the treatment group in each baseline quarter.

In I7, the treatment group dropped to 9,139 beneficiaries assigned to 16 practices, and the comparison group dropped to 27,717 beneficiaries assigned to 54 practices. This drop reflects the fact that only cohort 1 practices progressed to I7 and I8 at the time of this analysis. A similar phenomenon occurred with high-risk beneficiaries attributed to treatment and comparison practices, which dropped to 1,801 beneficiaries in the treatment group (down from 3,043 beneficiaries in the sixth intervention quarter) and 5,243 beneficiaries in the comparison group (down from 6,525 beneficiaries in I6; Table II.C.4).

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

The mean outcomes are calculated among all beneficiaries in the treatment and comparison group each quarter (presented in Tables II.C.3 and II.C.4). The means for the full patient population in Table II.C.3 can differ from the mean characteristics of the treatment and comparison practices (presented in the Table II.C.2). This happens because Table II.C.2 shows practice-level (not beneficiary-level) means, and larger practices do not receive larger weights. In contrast, in the beneficiary-level means presented in this section, larger practices will be represented more frequently than smaller practices.

Quality-of-care outcomes. For both the treatment and comparison groups during all baseline quarters, the number of hospitalizations for ACSCs among the full patient population was greater than the national average of 11.8 per 1,000 patients per quarter, and the number of hospitalizations for ACSCs among high-risk patients was generally more than three times the national average. Hospitalizations for ACSCs among all patients were 3 to 26 percent higher for the treatment group than the comparison group in all intervention quarters, with a consistent trend of increasing treatment-comparison differences from I2 to I6. A similar phenomenon occurred for high-risk beneficiaries. The 30-day unplanned readmission rates (number per q

Table II.C.3. Sample sizes and unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and comparison groups for FLHSA, by quarter

	Number of Medicare FFS beneficiaries (practices)			Inpatient admissions for ACSCs (#/1,000/quarter)			30-day unplanned hospital readmission rate (#/1,000/quarter)			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/month)		
Q	T	C (no wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Medicare FFS average																		
		49 million			11.8	n.a.		NA	n.a.		74	n.a.		105	n.a.		$860	n.a.
Baseline period (January 2012 to December 2013 for cohort 1 and June 2012 to May 2013 for cohort 2)																		
B1	13,337 (38)	31,180 (77)	13,288	16.3	15.7	0.7 (4.2%)	13.4	10.5	2.9 (28.1%)	85.5	76.6	8.9 (11.6%)	148.5	135.2	13.3 (9.8%)	$716	$690	$26 (3.8%)
B2	13,739 (38)	32,133 (77)	13,730	15.6	18.1	-2.5 (-14.0%)	13.6	10.7	3.0 (27.7%)	85.9	80.2	5.7 (7.1%)	162.6	135.3	27.3 (20.2%)	$778	$726	$52 (7.2%)
B3	14,080 (38)	32,992 (77)	14,079	13.4	15.4	-2.0 (-13.1%)	13.5	11.9	1.6 (13.7%)	78.6	79.3	-0.8 (-1.0%)	157.2	137.1	20.1 (14.6%)	$729	$699	$30 (4.3%)
B4	14,205 (38)	33,575 (77)	14,264	17.4	19.0	-1.6 (-8.6%)	13.6	12.7	0.9 (6.7%)	88.7	86.0	2.7 (3.1%)	150.6	146.1	4.5 (3.1%)	$792	$736	$56 (7.6%)
Intervention period (January 2013 to December 2014 for cohort 1 and June 2013 to December 2014 for cohort 2)																		
I1	13,391 (38)	30,633 (77)	13,011	18.4	14.6	3.8 (25.9%)	12.6	12.1	0.5 (4.4%)	83.3	79.1	4.1 (5.2%)	144.3	138.3	6.0 (4.3%)	$765	$736	$29 (3.9%)
I2	13,884 (38)	31,511 (77)	13,389	15.8	15.4	0.4 (2.5%)	12.2	10.6	1.5 (14.5%)	85.1	81.4	3.6 (4.5%)	158.4	144.4	14.0 (9.7%)	$809	$759	$51 (6.7%)
I3	14,262 (38)	32,034 (77)	13,591	14.4	13.4	0.9 (7.0%)	13.9	9.7	4.2 (43.8%)	84.5	71.8	12.6 (17.6%)	159.2	135.2	24.1 (17.8%)	$773	$743	$30 (4.0%)
I4	14,434 (38)	32,575 (77)	13,991	15.3	14.4	0.9 (6.4%)	14.0	11.2	2.8 (25.5%)	82.6	76.7	5.8 (7.6%)	168.7	134.4	34.3 (25.5%)	$812	$743	$69 (9.3%)
I5	14,500 (38)	32,448 (77)	14,308	17.9	15.6	2.3 (14.7%)	14.1	10.8	3.3 (30.3%)	85.9	83.9	2.1 (2.5%)	163.3	142.6	20.6 (14.5%)	$819	$753	$66 (8.8%)
I6	14,777 (38)	32,926 (77)	14,970	15.7	14.1	1.6 (11.0%)	12.7	11.2	1.5 (13.8%)	80.1	76.0	4.0 (5.3%)	165.3	145.2	20.1 (13.8%)	$814	$803	$12 (1.4%)
I7	9,139 (16)	27,717 (54)	8,975	16.5	15.6	0.9 (5.7%)	16.4	13.0	3.4 (26.6%)	92.2	83.0	9.2 (11.1%)	194.0	160.0	34.1 (21.3%)	$891	$847	$44 (5.2%)
I8	9,263 (16)	28,194 (54)	9,115	16.9	16.0	1.0 (6.1%)	13.4	10.6	2.8 (26.1%)	87.0	75.5	11.5 (15.3%)	186.6	155.5	31.2 (20.1%)	$800	$777	$23 (3.0%)

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. See Table II.C.2 for sources for the Medicare FFS averages.

Note: The baseline quarters are measured relative to the start of the baseline period in January 2012 for cohort 1 and June 2012 for cohort 2. For example, the first baseline quarter (B1) for cohort 1 runs from January to March 2012 for cohort 1 and from June to August 2012 for cohort 2. The intervention quarters are measured relative to the start of the intervention period. For example, the first intervention quarter (I1) runs from January to March 2013 for cohort 1 and June to August 2013 for cohort 2. In each period (baseline or intervention), the treatment group each quarter includes all Medicare FFS beneficiaries assigned to a treatment practice by the start of the quarter and

Table II.C.3 *(continued)*

who met other sample criteria—that is, they were enrolled in Medicare FFS and living in New York or Pennsylvania by the end of the relevant period. In each period, the comparison group each quarter includes all Medicare FFS beneficiaries assigned to a comparison practice by the start of the quarter and who met the other sample criteria. See text for details.

The outcome means were weighted, such that (1) each treatment beneficiary received a weight of 1; and (2) each comparison beneficiary received a weight that was the product of two weights: (a) a matching weight, equal to the reciprocal of the total number of comparison practices matched to the same treatment practice as the beneficiary's assigned practice, and (b) a practice size weight. The practice size weight differs depending on whether the beneficiary is at high risk or not. For high-risk beneficiaries, the weight equals the average number of high-risk beneficiaries assigned to the matched treatment practice during the four baseline quarters divided by the average number of high-risk beneficiaries assigned to the beneficiary's comparison practice over those quarters. For beneficiaries who were not at high risk, the weight equals the average number of non-high-risk beneficiaries assigned to the matched treatment practice during the four baseline quarters divided by the average number of non-high-risk beneficiaries assigned to the beneficiary's comparison practice over those quarters. The difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; Diff = difference; I = intervention; T = treatment; no wgt = unweighted; wgt = weighted.

NA = not available.

n.a. = not applicable.

Table II.C.4. Sample sizes and unadjusted mean outcomes for high-risk Medicare FFS beneficiaries in the treatment and comparison groups for FLHSA, by quarter

	Number of Medicare FFS beneficiaries (practices)			Inpatient admissions for ACSCs (#/1,000/quarter)			30-day unplanned hospital readmission rate (#/1,000/quarter)			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/month)		
Q	T	C (no wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Medicare FFS average																		
		49 million			11.8	n.a.		NA	n.a.		74	n.a.		105	n.a.		$860	n.a.
Baseline period (January 2012 to December 2013 for cohort 1 and June 2012 to May 2013 for cohort 2)																		
B1	3,374 (38)	7,302 (77)	3,341	43.6	43.9	-0.3 (-0.7%)	40.9	29.7	11.2 (37.6%)	201.2	179.8	21.5 (11.9%)	239.0	216.4	22.6 (10.5%)	$1,551	$1,536	$15 (1.0%)
B2	3,329 (38)	7,223 (77)	3,308	44.8	40.8	4.0 (9.7%)	34.8	26.4	8.4 (31.8%)	190.4	173.0	17.5 (10.1%)	268.2	222.5	45.8 (20.6%)	$1,632	$1,525	$107 (7.0%)
B3	3,269 (38)	7,118 (77)	3,259	33.0	40.3	-7.2 (-18.0%)	37.6	25.7	11.9 (46.3%)	175.9	161.1	14.8 (9.2%)	251.1	233.0	18.2 (7.8%)	$1,517	$1,361	$156 (11.5 %)
B4	3,153 (38)	7,005 (77)	3,217	48.8	51.1	-2.3 (-4.5%)	32.7	37.7	-5.0 (-13.3%)	188.4	181.5	6.9 (3.8%)	241.2	219.8	21.4 (9.7%)	$1,558	$1,529	$29 (1.9%)
Intervention period (January 2013 to December 2014 for cohort 1 and June 2013 to December for cohort 2)																		
I1	3,421 (38)	7,323 (77)	3,430	51.7	43.0	8.7 (20.2%)	37.1	37.4	-0.3 (-0.7%)	194.1	179.9	14.2 (7.9%)	226.2	231.4	-5.2 (-2.3%)	$1,644	$1,597	$48 (3.0%)
I2	3,363 (38)	7,243 (77)	3,389	40.4	38.1	2.4 (6.2%)	30.9	24.1	6.8 (28.1%)	188.2	168.6	19.7 (11.7%)	250.7	241.4	9.3 (3.8%)	$1,677	$1,445	$232 (16.1 %)
I3	3,300 (38)	7,066 (77)	3,292	38.2	33.9	4.3 (12.7%)	37.3	26.7	10.5 (39.4%)	188.8	154.7	34.1 (22.0%)	251.5	215.7	35.8 (16.6%)	$1,621	$1,503	$119 (7.9%)
I4	3,209 (38)	6,943 (77)	3,329	39.0	38.2	0.7 (1.9%)	38.3	30.3	8.0 (26.6%)	190.7	178.8	11.9 (6.7%)	295.1	213.9	81.2 (38.0%)	$1,695	$1,563	$132 (8.4%)
I5	3,126 (38)	6,693 (77)	3,284	47.0	36.6	10.5 (28.7%)	42.2	28.1	14.2 (50.5%)	200.6	197.2	3.4 (1.7%)	254.7	229.9	24.7 (10.8%)	$1,804	$1,625	$178 (11.0 %)
I6	3,043 (38)	6,525 (77)	3,325	38.1	35.6	2.6 (7.2%)	31.2	31.4	-0.2 (-0.6%)	166.0	168.0	-2.1 (-1.2%)	239.9	247.4	-7.5 (-3.0%)	$1,608	$1,554	$53 (3.4%)
I7	1,801 (16)	5,243 (54)	1,863	48.9	42.5	6.4 (15.0%)	49.4	33.3	16.1 (48.3%)	204.9	177.7	27.2 (15.3%)	297.1	259.4	37.6 (14.5%)	$1,813	$1,645	$169 (10.3 %)
I8	1,751 (16)	5,144 (54)	1,823	42.8	42.1	0.8 (1.8%)	37.1	26.7	10.5 (39.2%)	198.2	162.3	35.9 (22.1%)	281.6	257.0	24.6 (9.6%)	$1,605	$1,478	$127 (8.6%)

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. See Table II.C.2 for sources for the Medicare FFS averages.

Table II.C.4 *(continued)*

Note: See notes to Table II.C.3 for definitions of the baseline and intervention quarters. In each period (baseline or intervention), the treatment group each quarter includes all high-risk Medicare FFS beneficiaries assigned to a treatment practice by the start of the quarter and who met other sample criteria (also described in the note to Table II.C.3). In each period, the comparison group each quarter includes all Medicare FFS beneficiaries assigned to a comparison practice by the start of the quarter and who met the other sample criteria. See text for details. The outcome means were weighted, such that (1) each treatment beneficiary received a weight of 1; and (2) each comparison beneficiary received a weight that was the product of two weights: (a) a matching weight, equal to the reciprocal of the total number of comparison practices matched to the same treatment practice as the beneficiary's assigned practice, and (b) a practice size weight, which equals the average number of high-risk beneficiaries assigned to the matched treatment practice during the four baseline quarters divided by the average number of high-risk beneficiaries assigned to the beneficiary's comparison practice over those quarters. The difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; Diff = difference; I = intervention; T = treatment; no wgt = unweighted; wgt = weighted.

NA = not available.

n.a. = not applicable.

uarter) were generally higher for treatment practices than for comparison practices during baseline and intervention quarters—for both the full patient population and high-risk beneficiaries—with no notable trends in treatment–comparison differences during the intervention period.

Service use. During the baseline quarters, hospitalization rates for both the treatment and comparison groups were higher than the national benchmark of 74 per 1,000 patients per quarter (average of 85 and 81 for the full patient population in treatment and comparison groups, respectively), and over double the national benchmark for high-risk beneficiaries (average of 189 and 174 in treatment and comparison groups, respectively). Hospitalization rates among all patients were 3 to 18 percent higher for the treatment group than the comparison group in all intervention quarters, without any consistent trend of increasing or decreasing differences. A similar phenomenon occurred for high-risk beneficiaries.

At baseline, the ED visit rate among treatment and comparison practices was substantially higher than the national average of 105 per 1,000 patients per quarter among the full population (155 and 138 for the treatment and comparison groups, respectively), and more than double the national average among the high-risk population (250 and 223 for the treatment and comparison groups, respectively). ED rates among all patients were 4 to 26 percent higher for the treatment group than the comparison groups in all intervention quarters, without any consistent trend of increasing or decreasing treatment–comparison differences. A similar phenomenon occurred for high-risk beneficiaries, although there was more volatility from quarter to quarter in the ED rate among high-risk beneficiaries in the treatment group than in the comparison group.

Spending. For the full patient population in the treatment and comparison groups, mean monthly Medicare Part A and B spending was less than the national average of $860 per patient per month in all baseline quarters ($754 and $713 for the treatment and comparison groups, respectively); however, high-risk patients had significantly higher spending at baseline ($1,565 and $1,488 for the treatment and comparison groups, respectively). For the full patient population, the mean spending for the comparison group was within 9 percent of the treatment group during all intervention quarters; for high-risk patients, spending for the comparison group was within 16 percent of the treatment group during all intervention quarters. Similar to other outcomes, there was no consistent trend of increasing or decreasing treatment–comparison differences in spending during the intervention period, either among the full patient population or among high-risk patients.

c. Results for primary tests, by domain

Overview. The primary test results reflect the average impact estimate in the second and third year of the FLHSA intervention. For this report, we had data available for only the first 12 months of this period for cohort 1, and the first 6 months of this period for cohort 2. Thus, the primary tests in this report reflect impacts over only four intervention quarters for cohort 1 (I5 through I8) and only two intervention quarters for cohort 2 (I5 and I6). The impact estimates presented in Table II.C.5 reflect the average impacts across the two cohorts during these intervention quarters.

Table II.C.5. Results of primary tests for FLHSA

Primary test definition					Statistical power to detect an effect that is[a]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual)[b]	Size of the substantive threshold	Twice the size of the substantive threshold[b,c]	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual (standard error)[b]	Percentage difference[d]	*p*-value[e]
Quality of care outcomes (4)	Inpatient admissions for ACSCs (#/1,000 beneficiaries/ quarter)	Average over intervention quarters 5–8 for cohort 1 (January 1, 2014, to December 31, 2014) and quarters 5–6 for cohort 2 (July 1, 2014, to December 31, 2014)	All observable Medicare FFS beneficiaries attributed to treatment practices	-5.0%	21.4	38.1	16.8	2.2 (1.5)	15.1%	0.820[g]
	30-day unplanned readmissions (#/1,000 beneficiaries/ quarter)			-5.0%	20.9	36.7	14.1	0.1 (1.5)	0.9%	0.500[g]
	Inpatient admissions for ACSCs (#/1,000 beneficiaries/quarter)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices	-15.0%	42.7	81.9	44.2	5.0 (5.4)	12.9%	0.641[g]
	30-day unplanned readmissions (#/1,000/quarter)			-15.0%	40.2	78.4	40.0	1.0 (5.7)	2.5%	0.500[g]
	Combined (%)		All observable Medicare FFS beneficiaries	-10.0%	37.6	74.3	n.a.	n.a.	7.8%	0.776[h]
Service use (4)	All-cause inpatient admissions (#/1,000 beneficiaries/ quarter)		All observable Medicare FFS beneficiaries attributed to treatment practices[f]	-2.7%	24.6	46.3	86.3	1.1 (3.9)	1.3%	0.503[g]
	Outpatient ED visits (#/1,000 beneficiaries/ quarter)			-5.0%	49.2	89.3	177.3	1.7 (7.0)	1.0%	0.501[g]
	All-cause inpatient admissions (#/1,000 beneficiaries/ quarter)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices[f]	-4.9%	29.0	57.0	192.4	-5.4 (13.3)	-2.7%	0.490[g]
	Outpatient ED visits (#/1,000 beneficiaries/ quarter)			-15.0%	67.0	98.5	268.3	-17.8 (24.9)	-6.2%	0.438[g]
	Combined (%)		All observable Medicare FFS beneficiaries[f]	-6.9%	63.9	97.7	n.a.	n.a.	-1.7%	0.345[h]

Table II.C.5 *(continued)*

Primary test definition					Statistical power to detect an effect that is[a]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual)[b]	Size of the substantive threshold	Twice the size of the substantive threshold[b,c]	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual (standard error)[b]	Percentage difference[d]	*p*-value[e]
Spending (2)	Medicare Part A and B spending ($/beneficiary/ month)		All observable Medicare FFS beneficiaries attributed to treatment practices[f]	-1.6%	21.2	37.5	$831	-11.0 (27.9)	-1.3%	0.438[g]
	Medicare Part A and B spending ($/beneficiary/ month)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices[f]	-3.3%	24.6	46.3	$1,707	-5.9 (95)	-0.3%	0.498[g]
	Combined (%)		All observable Medicare FFS beneficiaries[f]	-2.5%	24.8	46.8	n.a.	n.a.	-0.8%	0.420[h]

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] The power calculation is based on actual standard errors from the analysis. For example, in the second-to-last row, a 3.3 percent effect on Medicare Part A and B spending would be a change of $56. Given the standard error of $95 from the regression model, we would be able to detect a statistically significant result 24.6 percent of the time if the impact was truly $56, assuming a one-sided statistical test at the $p = 0.10$ significance level.

[b] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[c] We show statistical power to detect a very large effect (twice the size of the substantive threshold) because this provides additional information about the likelihood that we will find effects if the program is indeed effective. If power to detect effects is less than 75 percent for even a very large effect, then the evaluation is extremely poorly powered for that outcome.

[d] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison groups, divided by the adjusted comparison group mean.

[e] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to 0 (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test.

[f] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

[g] We adjusted the *p*-values from the primary test results for the multiple (two) comparisons made within the service use domain, and (separately) for the two comparisons made within the quality-of-care outcomes domain.

[h] This *p*-value tests the null hypothesis that the difference-in-differences estimates across the multiple outcomes in the domain, each expressed as percentage change from the estimated counterfactual, are greater than or equal to zero (a one-sided test).

ACSC = ambulatory care-sensitive condition; ED = emergency department; FFS = fee-for-service; FLHSA = Finger Lakes Health Systems Agency; HCIA = Health Care Innovation Award.

n.a. = not applicable.

We conducted 2 primary tests for each of the five outcomes discussed next: one for all patients and one for high-risk patients, for a total of 10 primary tests. In all 10 tests, regression-adjusted differences between the treatment and comparison groups during the primary test period were statistically insignificant. All three domain-level tests, in which all outcomes in the same domain were combined, yielded statistically insignificant regression-adjusted differences that were smaller than the substantive thresholds in either direction.

Quality-of-care outcomes. The rate of ACSC hospitalizations for the treatment group during the primary test period was 15 and 13 percent higher than our estimate of the counterfactual for all patients and high-risk patients, respectively, and the rate of unplanned readmissions was 1 and 3 percent higher for all patients and high-risk patients, respectively. (Our estimate of the counterfactual is the treatment group mean minus the difference-in-differences estimate.) Although the treatment–comparison differences in ACSC hospitalizations are substantive, none of these differences were statistically significant. After combining results across the two outcomes (and both populations) in this domain, the outcomes for the treatment group were 8 percent higher than outcomes for the estimated counterfactual, and this difference was not statistically significant nor larger than the substantive threshold.

The statistical power to detect effects for ACSC hospitalizations and 30-day unplanned readmissions was poor. For example, Table II.C.5 indicates that the tests had a 21 percent likelihood of detecting an effect on ACSC hospitalizations among the full patient population that was the size of the substantive threshold. Power was 43 percent among the high-risk population, which is still poor, but better than the power for the population because the substantive threshold is larger for the high-risk population.

Service use. The treatment group's average hospitalization and outpatient ED visit rates were 1 percent higher than the estimate of the counterfactual for the full population, but 3 to 6 percent lower than the counterfactual for the high-risk population for these outcomes. None of these differences was statistically significant or substantively large. After combining results across the two outcomes (and both populations) in this domain, the outcomes for the treatment group were similar to the outcomes for the counterfactual, and the difference between the two was not statistically significant. For the full patient and high-risk patient populations, power to detect effects that were the size of the substantive thresholds was poor for hospitalization rates (ranging from 25 to 29 percent), and marginal for ED visits (ranging from 49 to 67 percent).

Spending. The full patient population in the treatment group averaged $831 per person per month in Part A and B spending during the primary test period, which was 1.3 percent (or $11) lower than the estimated counterfactual. This difference was not statistically significant. Similarly, high-risk beneficiaries in the treatment group had nearly identical Part A and B spending as the estimated counterfactual (high-risk beneficiaries in the comparison group) during the primary test period. Combining results across both populations in this domain, treatment group spending was 1.0 percent lower than the counterfactual, and the difference between the two was not statistically significant. Statistical power to detect an effect the size of the substantive threshold was poor (below 25.0 percent for both populations).

d. Results for secondary tests

As shown in Table II.C.6, the differences in nearly all outcomes for the treatment group and its estimated counterfactual were small and not statistically significant during the secondary test period: the first 12 months of the intervention (I1 through I4). One exception is a sizable difference for ACSC hospitalizations among the full patient population, in which the rate for treatment practices was 22 percent higher than the estimated counterfactual. However, this large difference is likely due to chance (and not a signal of a limitation in the comparison group), given that only 1 of 10 secondary tests found large differences and that ACSC hospitalization is a particularly volatile measure (as indicated by the very low statistical power to detect effects).

Overall, these results help support the credibility of the comparison group because we generally do not see large differences (favorable or unfavorable) during the first year of the intervention, a period during which we and the awardee did not expect to see large program effects. This increased confidence in the comparison group, in turn, gives us greater confidence in the primary test results.

e. Consistency of quantitative estimates with implementation findings

The impact estimates from the primary tests are plausible given the implementation findings. The primary tests did not find any effects (favorable or unfavorable) during the primary test period covered in this report (year 2 of the intervention) that were statistically significant or substantively important. The implementation evidence shows the program was active during these 12 months. For example, as described in Section II.A.2, as of December 2014, FLHSA care managers provided services (both intensive and otherwise) to 14,472 distinct patients—roughly equivalent to the sample size for our analysis of program impacts on the full patient population. (However, it should be noted that some portion of patients who received services are not Medicare FFS patients, and thus are not included in this analysis.) Therefore, the lack of measured effects is not simply due to the program failing to deliver a meaningful intervention. However, even with a well-implemented intervention, it is possible that the program was unable to change patients' or providers' behaviors in ways that would affect study outcomes during the primary test period covered in this report.

f. Conclusions about program impacts, by domain

Based on all evidence currently available, we have drawn the preliminary conclusion that the program impact **is indeterminate in each of the three domains**: quality-of-care outcomes, service use, and spending. These conclusions are summarized in Table II.C.7. We reached these conclusions because (1) in each domain, the primary test results were neither statistically significant nor substantively large; and (2) the secondary tests helped to confirm the credibility of the comparison group used in the primary tests by showing that there were no statistically significant effects in the first year of program operations—a period when we and the awardee expected few or no effects.

Table II.C.6. Results of secondary tests for FLHSA

Secondary test definition				Results			
Domain	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual (standard error)	Percentage difference[a]	*p*-value[b]
Quality-of-care outcomes	Inpatient admissions for ACSCs (#/1,000 beneficiaries/quarter)	Average over intervention quarters 1–4 (January 1, 2013, to December 31, 2013, for cohort 1 and July 1, 2013, to June 31, 2014, for cohort 2)	All observable Medicare FFS beneficiaries attributed to treatment practices[c]	16.0	2.9 (1.4)	22.3%	0.982
	30-day unplanned readmissions (#/1,000 beneficiaries/quarter)			13.2	0.1 (1.3)	1.1%	0.542
	Inpatient admissions for ACSCs (#/1,000 beneficiaries/quarter)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices[c]	42.3	5.2 (4.6)	13.9%	0.871
	30-day unplanned readmissions (#/1,000 beneficiaries/quarter)			35.9	-1.8 (4.8)	-4.8%	0.353
Service use	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)		All observable Medicare FFS beneficiaries attributed to treatment practices[c]	83.9	2.9 (3.6)	3.6%	0.788
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			157.6	1.1 (6.2)	0.7%	0.572
	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices[c]	190.5	2.9 (11.5)	1.6%	0.600
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			255.9	1.6 (23.7)	0.6%	0.528
Spending	Medicare Part A and B spending ($/beneficiary/month)		All observable Medicare FFS beneficiaries attributed to treatment practices[c]	$790	3.0 (24.6)	0.4%	0.549
	Medicare Part A and B spending ($/beneficiary/month)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices[c]	$1,659	25.8 (81.1)	1.6%	0.625

Table II.C.6 *(continued)*

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison groups, divided by the adjusted comparison group mean.

[b] The *p*-values from the secondary test results were *not* adjusted for multiple comparisons within each domain or across domains.

[c] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

ACSC = ambulatory care-sensitive condition; FFS = fee-for-service; FLHSA = Finger Lakes Health Systems Agency.

Table II.C.7. Preliminary conclusions about the impacts of FLHSA's program on patients' outcomes, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion		
		Primary test result(s) that supported conclusion	Primary test result plausible given secondary tests?	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	Indeterminate effect	No individual test in the domain was statistically significant The combined test across both outcomes in the domain was not statistically significant or substantively important	Yes	Yes
Service use	Indeterminate effect	Same as above	Yes	Yes
Spending	Indeterminate effect	Same as above	Yes	Yes

Sources: Tables II.C.5 and II.C.6.

FLHSA = Finger Lakes Health Systems Agency.

These indeterminate effects in each of the three domains have two possible interpretations. First, the program might not have an effect in any of the domains for the population and period covered in this report. Alternatively, the program could have had an effect in one or more of the domains—and possibly even one that exceeded the substantive thresholds—but, due to the statistical uncertainty in the estimates, we were unable to detect it.

As mentioned earlier, these conclusions are preliminary because the analyses do not yet cover the full period that we will include in the final impact analysis in future reports, nor do they include cohort 3 practices. It is possible that, when we extend the final evaluation to include eight additional implementation quarters, up to 20 additional practices, and Medicaid data, the program will have measurable effects in one or more of the domains. In addition, the statistical power to detect effects on these outcomes will likely increase as the analysis incorporates additional cohort 3 practices and includes additional implementation quarters.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

FLHSA received HCIA funding to help participating practices become PCMHs, to implement intensive care management of high-risk patients, and to work with two area insurers to develop a communitywide outcomes-based payment model. The program aims to improve the quality of care for high-risk Medicaid and Medicare beneficiaries while reducing the cost of care. The program has been implemented largely as intended, aided by (1) the ability of individual practices to adapt the program to their needs and (2) the support of highly engaged practice staff led by practice champions and managers. Participating staff felt the program improved care delivery by focusing on team-based care, care management, and patient communication.

Program implementation was also facilitated by practices' use of data to monitor performance and conduct quality improvement activities, strong care team communication and collaboration within practices, and ACO support for practice transformation. Implementation was hindered by challenges in integrating CHWs into practices, competing demands on practice staff time, and streamlining program and ACO requirements for care management. The HCIA Primary Care Redesign Clinician Survey found that most clinicians believed that the HCIA-funded initiative would have a positive effect on the patient-centeredness, quality, and timeliness of care and patients' safety.

Preliminary results from the impact evaluation found no measurable effects of the program on quality-of-care outcomes (30-day readmissions or hospitalizations for ACSCs), service use (all-cause hospitalizations or outpatient ED visits), or Medicare Part A and B spending for either the full Medicare FFS patient population or high-risk Medicare FFS beneficiaries during the first 12 months of the primary test period (months 12 through 24 after the program began). For all three domains (quality of care, service use, and spending), the lack of measured effects might be because the program did not have effects or that it did but, due to poor statistical power, our tests failed to detect them. The program could have measurable impacts in one or more of three domains when the evaluation is extended to (1) cover the full primary test period (months 12 through 42 after the program began); (2) add Medicare beneficiaries served by the cohort 3 practices; and (3) if data permit, add Medicaid beneficiaries served by the treatment practices.

Our next steps for this evaluation are to (1) monitor FLHSA's program implementation and any plans for sustaining the program through the end of its award in June 2016; (2) evaluate trainees' and clinicians' attitudes and experiences with the program in the third year of the award through administered surveys; (3) extend the impact evaluation to include the full time period of program operations, the third cohort of program participants, and Medicaid data; and (4) use the implementation findings to help interpret the impact results.

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

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table A.1. Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no. 9, 2014, pp. 1672–1679.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries (mean)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Health Indicators Warehouse. "Medicare Beneficiaries Eligible for Medicaid (percent)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014c. Available at http://www.healthindicators.gov/Indicators/Medicare-beneficiaries-eligible-for-Medicaid-percent_317/Profile/ClassicData. Accessed August 4, 2015.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Langley, G.L., R. Moen, K.M. Nolan, T.W. Nolan, C.L. Norman, and L.P. Provost. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance* (2nd edition). San Francisco: Jossey-Bass Publishers; 2009.

Medicare Payment Advisory Commission. "Beneficiaries Dually Eligible for Medicare and Medicaid." Washington, DC: MEDPAC, December 2013.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21.

Truven Health Analytics. "AHRQ Quality Indicators, Prevention Quality Indicators v5.0 Benchmark Data Tables." Prepared for the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. Santa Barbara, CA: Truven Health Analytics, March 2015. Available at http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V50/Version_50_Benchmark_Tables_PQI.pdf. Accessed August 18, 2015.

Van Walraven, C., I.A. Dhalla, C. Bell, E. Etchells, I.G. Stiell, K. Zarnke, P.C. Austin, and A.J. Forster. "Derivation and Validation of an Index to Predict Early Death or Unplanned Readmission After Discharge from Hospital to the Community." *Canadian Medical Association Journal,* vol. 182, no. 6, April 6, 2010, pp. 551–557.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Foundation for California Community Colleges and the Transitions Clinic Network

March 2016

Rachel Shapiro
Jennifer Lyons
Boyd Gilman
Catherine DesRoches
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day

Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno

Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

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

FOUNDATION FOR CALIFORNIA COMMUNITY COLLEGES AND THE TRANSITIONS CLINIC NETWORK

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by the Transitions Clinic Network (TCN) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the TCN program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). We were unable to conduct an impact evaluation of the TCN program because we could not construct a valid comparison group for the awardee's unique and vulnerable treatment population or measure outcomes for the treatment group before program implementation. As a result, we do not provide an estimate of the impact of the program on patients' outcomes in this report. In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF TCN

TCN received a three-year, $6.8 million HCIA to provide high quality care to formerly incarcerated people at 13 clinics with specialized programs located in community health centers in six states (Alabama, California, Connecticut, Maryland, Massachusetts, and New York) and Puerto Rico (see Table I.1). The TCN program, an expansion of an existing care transition model for people recently released from prison, is administered by three partners—the University of California-San Francisco, the City College of San Francisco (CCSF), and Yale University. The Foundation for California Community Colleges is the program's fiscal agent; the foundation has no role in program implementation. The goal of the TCN program is to improve patients' clinical outcomes, self-reported health status, and satisfaction with their care, and to lower patients' health care costs by reducing unnecessary hospital admissions, emergency department (ED) visits, and duplicated diagnostic tests. The program was scheduled to end on June 30, 2015, but TCN received a no-cost extension for 12 months to support continued clinic operations and data collection on patients enrolled in the program. The program end date is now June 30, 2016.

Table I.1. Summary of TCN PCR program

Awardee name	Foundation for California Community Colleges and the Transitions Clinic Network
Award amount	$6,852,153
Implementation date	August 2012
Award end date	June 30, 2016
Program description	1. Provide and coordinate primary care and other health and social services for chronically ill patients recently released from prison 2. Improve quality of care provided at participating clinics through collection and use of patients' data using a cloud-based data platform 3. Expand the reach of the Post-Prison Health Worker training program for formerly incarcerated community health workers (CHWs)
Innovation components	Care coordination, patient navigation, care management
Intervention focus	Patient

Table I.1 *(continued)*

Workforce development	Hire new staff or retrain existing staff to serve as CHWs; provide training to CHWs, liaisons, and other clinical staff
Target population	Adults released from prison within the past six months, enrolled in Medicaid or Medicaid-eligible, and with chronic health conditions or older than 50
Program setting	Provider-based (13 clinics based in hospitals and community health centers)
Market area	Multistate (Alabama, California, Connecticut, Maryland, Massachusetts, and New York), and Puerto Rico
Market location	Urban (all)
Core outcomes	• Improvement in quality of and access to care, clinical outcomes, patients' satisfaction, and self-reported health status • Reduction in cost of care through decreased unnecessary hospital admissions, ED visits, and duplicated diagnostic tests

Source: Review of TCN program reports, March 2015.

Note: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health IT systems, and other operational activities.

II. SUMMARY OF FINDINGS

A. Program Implementation

In this section, we first provide a detailed description of the intervention, highlighting how it has been adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of the implementation of the TCN program on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in May 2014 and March 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

The TCN program includes one main structural component (see Table II.A.1). To provide and coordinate on-site primary care and other health and social services for patients recently released from prison, TCN trained people with a history of incarceration to work as community health workers (CHWs) and integrated them into existing primary care teams at 13 participating clinics housed in Federally Qualified Health Centers (FQHCs) or other community health centers located in high-need communities affected by incarceration. TCN leadership recruited clinics based on existing relationships between TCN and clinic staff, staff's willingness to implement the program model, and clinics' ability to accommodate additional staff (CHWs) and an increased patient population. (Five clinics already were implementing some aspects of the TCN program model, and TCN leadership assessed the remaining eight clinics as having the necessary

resources and capacity to implement it.) The local evaluation team at Yale University worked with clinic staff to collect patient data using a cloud-based data platform; the Yale evaluation team will continue to provide this support during the no-cost extension period. The TCN program uses these data to provide participating clinics with real-time information on patients' characteristics—such as housing and insurance status and whether they experienced relapse to illicit drug use—and services patients receive (for example, clinic services, ED visits, and hospital admissions). By providing these data, TCN leadership hope to improve clinic staff's understanding of patients' backgrounds and service use and the quality of care the clinics provide. The Yale evaluation team is also using practice-reported data to evaluate the program's effect on patients' use of health care and other services and estimate the associated impact on cost of care for the broader TCN patient population.

Table II.A.1. Key details about program design and adaptation

	Program component
	Providing and coordinating primary care and other services
Target population	• Adults in 13 participating clinics' service areas who were released from prison in the past six months, have Medicaid or are Medicaid-eligible, and either have chronic health conditions related to behavioral health, substance use, or physical health or are older than 50.
Patient recruitment and enrollment	• In most clinics, CHWs identified potential patients through outreach to parole and probation departments, local hospitals, and community organizations serving people recently released from prison. • Some clinics also received patient referrals directly from prisons or jails. • When permitted, CHWs reached out to potential patients before their release from prison. • After identifying a potentially eligible individual, clinic staff (CHW, panel manager, or other clinic staff) used an online survey tool to collect the patient's demographic information and determine whether the patient met participation criteria. During this process, staff also asked patients for consent to enter their data into the online data platform during their participation in the program.
Service delivery protocol	• The TCN program aims to engage patients in primary care within one month of release from prison through a primary care provider visit and a meeting with a CHW. • TCN recommends that CHWs initially contact patients weekly or biweekly, and adjust contact based on patients' needs. • TCN expects CHWs to follow up within one week with any patients who have an acute care visit. • CHWs document each patient encounter. • CHW or other clinics staff conduct baseline and follow-up surveys in six-month intervals for patients who consent to having their data collected in the online platform. Those who meet criteria may decline to have their data collected in the online platform and still receive program services. • TCN participants continue in the program for as long as they choose to engage in services.
Adaptations	• One clinic began using a risk-stratification protocol in November 2014 to help CHWs prioritize their time such that most time is spent with patients recently released from prison and disconnected from needed services and care.

Sources: Interviews from second site visit, March 2015; document review, March 2015.

b. Target populations and patient identification, recruitment, and enrollment

As described in Table II.A.1, the TCN program targets people released from prison in the previous six months who have chronic health conditions or who are older than 50. These people are often released from prison without connections to primary care or other health services and face many barriers to accessing care, such as homelessness and lack of familiarity with the health care system. Participating clinics used a variety of approaches to recruit patients. CHWs identified potential patients through outreach to parole and probation departments, local hospitals, community organizations serving people recently released from prison, and other venues. After identifying potential patients, the CHW, panel manager, or other clinic staff completed a demographic survey with the patients and enrolled them in the program.

c. Intervention staff and workforce development

Intervention staff. As shown in Table II.A.2, the TCN program model has three key staff roles: (1) CHWs, recruited because they have a history of incarceration; (2) primary care providers; and (3) liaisons to connect the clinic with the program leadership. The TCN program also provides funding to support a part-time panel manager at each clinic. Most clinics use this funding to support data collection by CHWs and other clinic staff members. However, some clinics hired a new staff person dedicated to these responsibilities. To implement the program, the 13 participating clinics hired new staff or retrained existing staff to serve as CHWs and integrated them into existing primary care teams. The types of staff on the primary care teams vary by clinic, based on the standard staffing structure at the organization in which the clinic is housed. In some clinics, providers work directly with the CHWs to coordinate patient care; in others, providers work with social workers or other intermediate staff who, in turn, engage the CHWs. Some clinics have formal weekly meetings for clinic staff, whereas others rely on CHWs and providers to communicate by telephone or messages in the electronic health record (EHR).

CHWs' responsibilities vary among the participating clinics based on their skill sets and the resources available at the organization housing the clinic. For example, some CHWs work with social workers who manage patients' referrals to social and community services; in other clinics, CHWs perform this case management role. Although the range of responsibilities varies, TCN program leadership expects all CHWs to carry out several core tasks, including (1) conducting outreach to new patients and follow up to existing patients, (2) advocating for patients and help them navigate the health and social service systems, and (3) providing mentoring and peer support. In response to some clinics' early challenges using CHWs effectively in these roles, in September 2014 TCN leadership introduced a CHW assessment tool that provides additional guidance on the CHW role. The tool specifies that CHWs should spend 40 percent of their time conducting outreach and providing services in community settings, with the rest of their time spent in the clinic. The tool also lists key skills that CHWs are expected to develop while employed at the clinic, such as case documentation and management, health care system navigation, patient education, medication reconciliation, data collection and reporting, and billing (when applicable). In addition, clinics are expected to provide CHWs with professional development opportunities and structured supervision and to include CHWs in team meetings and case conferences as fully integrated clinical team members. CHWs' caseloads vary across

Table II.A.2. Key details about TCN intervention staff

Program component	Staff members	Staff responsibilities	Adaptations
Providing and coordinating primary care and other services	CHWs	• Work with a clinic liaison, primary care providers, and other clinical staff to engage patients in on-site primary care and connect them to additional health and social services, such as housing and food resources. • Conduct outreach to potential patients. • Advocate for patients. • Provide mentoring and peer support to help patients navigate the health and social service systems. • Use the online data platform to track patient interactions. • Other responsibilities as appropriate based on CHWs' skill sets and the clinics' available resources.	In September 2014, TCN introduced a CHW assessment tool to help clinics integrate CHWs into care teams and provide them with appropriate and useful supervision, and to show them how to allocate their time among tasks.
	Primary care providers	• Provide primary care services to patients. • Work with CHW to coordinate care with other services.	None.
	Clinic liaisons	• Administrative or staff person at each clinic serves as a connection between the clinic and TCN. • Oversees CHWs.	None.
	Panel manager (CHWs or other clinic staff)	• Collects patient data using health surveys housed on the online data platform.	Some clinics hired additional staff to collect and manage these data; at other clinics, primary care providers or other clinic staff collected and managed these data.

Sources: Interviews from second site visit, March 2015; document review, March 2015.

the participating clinics. TCN leadership recommends that CHWs carry caseloads of 30 to 40 patients, reflecting CHWs' schedules and skill levels, as well as the patients' levels of need.

In addition to implementing the operational aspects of the initiative, TCN program partners developed and provided training to staff at participating clinics. First, in May 2014 CCSF converted its existing post-prison health worker certification program to an online format to train formerly incarcerated people to be CHWs at the participating clinics. The 20 units of interactive online courses emphasize practical skills to apply in the clinic setting (for example, motivational interviewing and management of chronic health conditions). The units also provide a conceptual understanding of the health impacts of incarceration (for example, the effect of solitary confinement on mental health). All CHWs at the participating clinics enrolled in, or had previously completed, the post-prison health worker certification program. CHWs were allotted five hours per week while at the clinic to dedicate to their coursework. Second, TCN program staff provided training and guidance on program implementation to clinic liaisons and other clinic staff through monthly meetings (with all clinics) and quarterly meetings (with individual

clinics). Third, the TCN program offered training on cultural competency and the transitions clinic model to staff. This training educated providers and other clinic staff on the unique needs of people recently released from prison. Finally, to foster connections among the workforce, the TCN program leadership hosted two annual in-person retreats (in October 2013 and September 2014) for clinic staff to encourage shared learning across clinics.

d. Service delivery protocols

The TCN program model does not require adherence to protocols for providing or coordinating patients' care. Although the TCN model offers general guidelines that clinics can use to help them with program implementation, it is intended to be adapted to the specific needs of the clinic, its staff, and its patients. During a February 2015 telephone interview, one program leader described this approach as "walking a fine line of helping to create structure, but being open to what may work at the local level and understanding that the program needs to have some flexibility because the sites are all structured differently, have different resources available, and have different staffing."

Most of TCN's structured guidance applies to the clinics' first interactions with a patient recently released from prison. For example, TCN instructs clinic staff to attempt to engage patients in primary care within one month of release from prison. During the first appointment, TCN staff (CHW, panel manager, or primary care provider) use an online survey to collect patients' demographic information. As part of this survey, the staff person determines whether the patient meets the criteria for the HCIA-funded program and records the eligibility status in the tool. If the patient meets the criteria and consents to data collection using the cloud-based platform, TCN staff complete a baseline survey with the patient and conduct follow-up surveys in six-month intervals. Also during the first appointment, the primary care provider addresses patients' urgent medical issues, ensures that patients receive medication refills, and offers ongoing primary care services to patients without an existing primary care provider. CHWs meet with patients before, during, or after this first visit to learn about their unique needs. This needs assessment extends beyond medical needs and includes a broad set of topics, such as housing, family reunification, health literacy, and substance use. These needs dictate the amount and types of services provided. In addition to meeting patients in the clinic, CHWs meet with patients in the community and conduct telephone follow-up.When enrolled in the TCN program, the amount of contact CHWs have with patients varies. TCN program administrators recommend that CHWs contact patients weekly or biweekly when patients first engage in the program, but suggest that CHWs adjust the amount and types of services based on each patient's needs. TCN leadership also expects CHWs to follow up within one week with any patients on their caseload who have an acute care visit. Those requiring more guidance, such as patients with serious medical or mental health conditions, might be in daily communication with the CHW.

CHWs are required to document each patient encounter. They record (1) the person who initiated the contact; (2) the type of communication (telephone, email, text, or face-to-face); (3) where and for how long the contact occurred; (4) the types of encounters that occurred (such as contacting the patient about a clinic visit, following up after an acute care visit, or accompanying the patient to an appointment at another location); (5) the issues that were addressed (medication, chronic disease self-management, health or social service care coordination or navigation, and

mental health or substance use); (6) the CHW's confidence that the patient can achieve the identified goals of the visit; and (7) the plan for following up with the patient. As patients become connected to other community resources and improve their ability to navigate the health and social service systems independently, they require less frequent contact by CHWs.

Working with TCN program leaders, in November 2014 staff at one clinic developed a protocol for stratifying patients based on their levels of risk. Under the risk-stratification protocol, CHWs classify each patient in one of three color-coded groups: (1) red (recently released from prison and disconnected from needed services and care), (2) yellow (engaged in services but still in need of support), or (3) green (mostly functioning independently and in need of minimal support). This protocol helps CHWs prioritize their time, recognizing that some patients will want more frequent contact despite having less acute needs.

Program leaders and staff at this clinic are also pursuing a formal process for transitioning patients out of the program when they no longer require ongoing CHW support. The TCN program does not have formal guidelines for graduating participants from the program. Staff recognize that some patients remain on CHW caseloads for several years despite having successfully navigated the transition from prison to the community. As one clinic staff noted during our second site visit, "Clients get very attached to [CHWs] and we need to be sure we're empowering patients to be more in charge of their own medical care." To facilitate the transition off the CHW caseload, program leaders and clinic staff developed a script that helps CHWs communicate to patients that the transition is a sign of progress and growth. At the time of the site visit, only one clinic was using the risk-stratification and transitions protocols; program leaders noted that other clinics will begin using these risk-stratification and transition protocols as caseloads increase and patients begin to reach the graduation point.

TCN program leaders also support the clinics' use of patient data to improve care. The Yale evaluation team processes data collected through the online surveys and provides clinics with access to these data in several formats. For example, TCN uses the online database to generate graphic summaries (dashboards) to identify at-risk patients who might benefit from targeted interventions. At-risk patients include uninsured patients, those who recently used the ED, and those with unstable housing. Clinics also have access to summaries of the individual patient's health risk behaviors generated after the patient completes the baseline and follow-up surveys. In addition, the evaluation team provides each site with monthly summaries of their patients' data compared with all participating clinics. Although these data are available to all participating clinics, there are no formal protocols for how the clinics should use them. Few interviewed clinic staff reported using these data. In a February 2015 telephone interview, one program administrator noted that "Each clinic has tried to figure out a way to integrate the types of information and data that we're providing into their clinical practice, and some have found it more useful than others."

2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness. For our assessment of implementation effectiveness, we rely on

telephone and site visit interviews with program administrators in February and March 2015, respectively, as well as self-reported information included in the quarterly self-monitoring and measurement reports.

a. Program enrollment

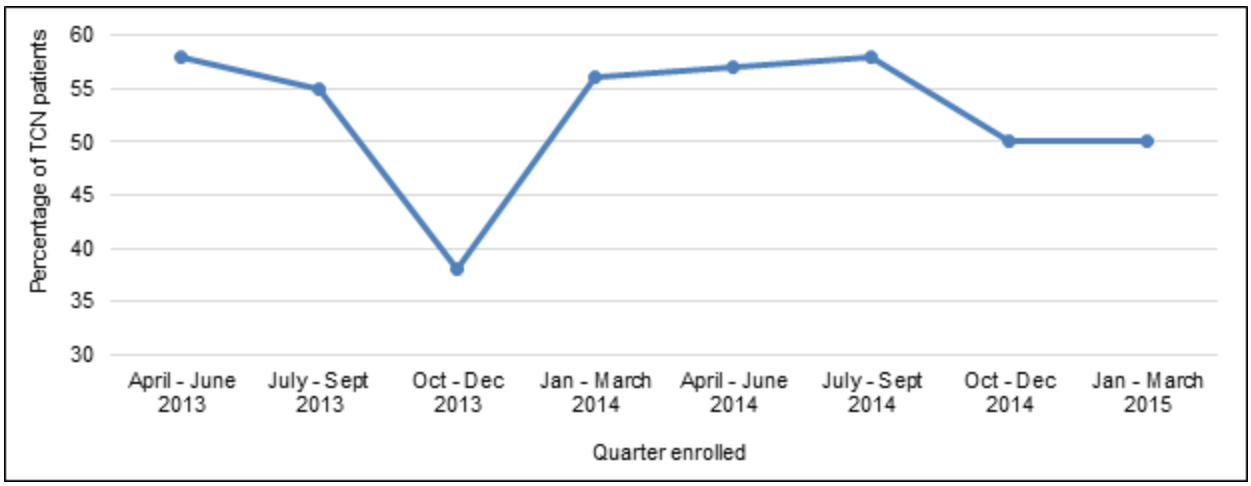
As of March 2015, the 13 participating clinics had served a total of 1,351 distinct patients, falling short of its target of 1,588 patients. Of the 1,351 distinct patients, 722 consented to have their information entered into the online data platform. Program leaders report that the program initially struggled to meet its enrollment targets as a result of several challenges, such as the length of the institutional review board approval process, unexpected changes in patient referral sources related to expanding health insurance coverage, and several clinics' issues with appropriately using and supporting CHWs to conduct patient outreach. In response to these early challenges, TCN program leadership worked with clinics to improve outreach strategies by developing targeted outreach plans and tracking referral sources. These strategies resulted in increased enrollment. However, TCN program leadership decided to stop enrolling new patients in March 2015 to focus its remaining resources on collecting follow-up data from enrolled patients for use in the program's internal evaluation (discussed in more detail in Section II.A.2.c).

b. Service measures

TCN's self-monitoring metrics indicate that clinic staff are reaching and retaining patients recently released from prison but continue to face barriers to engaging this population. The TCN model aims to engage patients within the first month of their release from prison and, over the course of the program, TCN clinics achieved this goal for more than half of the enrolled patients (Figure II.A.1). During a February 2015 telephone interview, TCN program leadership described the program's goal of engaging patients within one month as "well above the standard for how long it takes for individuals to get into care," noting that they intentionally set a goal they thought was "achievable, but also something to strive for."

However, according to the awardee's self-reported data, CHWs' success in contacting new patients during the first two weeks of their enrollment in the program declined over the course of the program, ranging from a high of more than 90 percent of patients from April to June 2013 to fewer than one-quarter of patients during January to March 2015 (Figure II.A.2). TCN program administrators were not entirely clear why the decline in engagement rates has occurred; some suggested it might not be a true reflection of engagement but rather a delay in data entry as CHW caseloads increased (that is, entering data for patients a few weeks after enrollment occurred). Indeed, we heard about significant barriers related to using the online data platform, although we also heard about significant barriers to engaging this population (discussed in Section II.A.3.d), which could also explain the reported decline in engagement rates. Because TCN did not report data on CHW caseloads, we were unable to assess the relationship between caseload size and engagement.

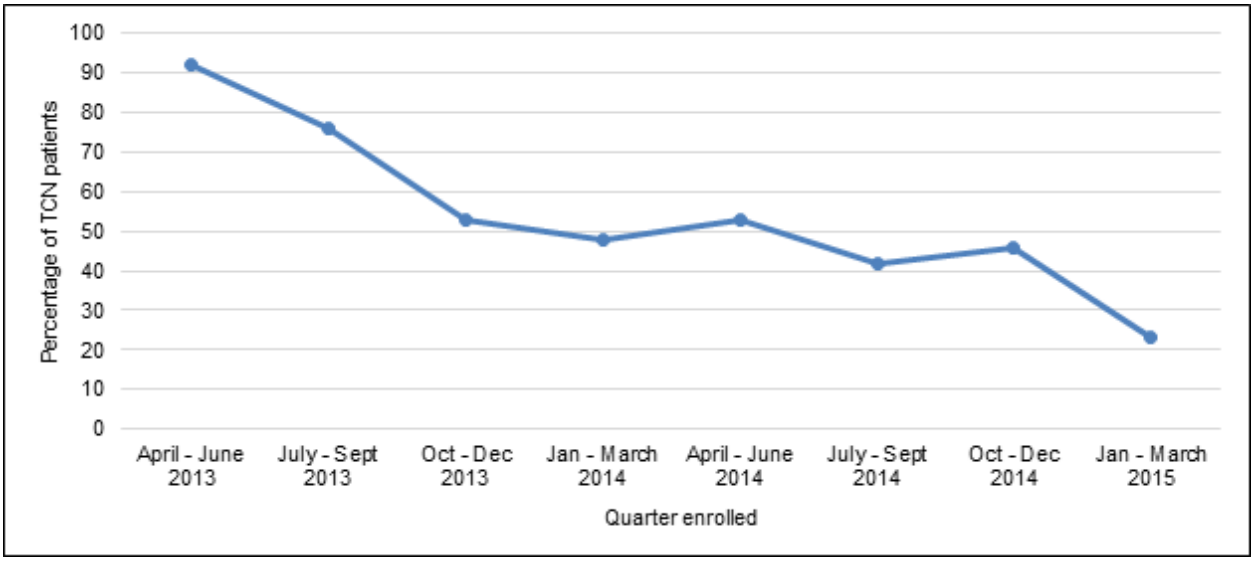
Figure II.A.1. TCN self-reported percentage of patients with a recorded primary care visit within one month of release from prison



Source: March 2015 TCN Measurement and Monitoring Plan.

Note: This information is based on the awardee's self-reported data. We have not attempted to verify its completeness or quality. The numerator is the number of patients with a recorded primary care visit within one month of release from prison. The denominator is all TCN patients during the measurement year. Of the 14 clinics, 11 collected data from April to September 2013. All 14 clinics collected data from October 2013 to March 2015.

Figure II.A.2. TCN self-reported percentage of patients with a recorded CHW encounter within first two weeks of program enrollment



Source: March 2015 TCN Measurement and Monitoring Plan.

Note: This information is based on the awardee's self-reported data. We have not attempted to verify its completeness or quality. The numerator is the number of patients seen who, within two weeks of enrollment, have a completed CHW encounter form. The denominator is the number of new TCN patients during the measurement year.

Although TCN self-reported data showed difficulty in initially engaging patients, these data also showed that the program engaged more than half of them in ongoing primary care. Of patients who had been in the program for at least six months, more than 50 percent had at least two visits with a TCN primary care physician in the past year. TCN program administrators reported that, although they did not set specific goals for this or other measures, they felt the 50 percent rate was a positive sign, given (1) the significant barriers to engaging this population and (2) program administrators' expectations for less frequent primary care visits for many patients after the first six months of the program, when they thought most patients would be better able to manage their chronic conditions.

c. Staffing measures

In addition to monitoring clinics' efforts to engage patients and provide and coordinate their care, TCN staff also tracked CHWs' training and retention at participating clinics. The 13 participating clinics hired or retrained existing staff as CHWs within the planned time line. All 10 CHWs from the original cohort of the TCN program completed the online post-prison health worker training program by May 2015. The 3 CHWs hired at the sites added later in the implementation period are completing the certificate program in person at a CCSF satellite campus. The TCN program retained all CHWs staffed at the participating clinics. One program leader noted during the second site visit that the program's success in retaining all CHWs "really speaks to the recruitment of the right CHWs and to the benefit of having this partnership that really allowed us to work closely with the [clinical program leaders] to create relevant curriculum and to give CHWs the support they need."

d. Program time line

The participating clinics hired or retrained CHWs according to schedule and the CHWs are helping patients engage in on-site primary care and connect to other health and social services as planned. To provide this training, CCSF adapted the post-prison health worker training program to an online format to provide training to CHWs at the participating clinics. All 20 units of the training program were developed on schedule and available online by May 2014.

However, TCN faced early delays in helping clinics collect and use patients' data to improve quality of care. Initially, TCN leadership hoped that the clinics would be able to start collecting data within a few months of program start-up. However, this proved infeasible for some clinics because of two issues: (1) delayed institutional review board approval and (2) frontline staff's difficulty collecting and inputting data (as discussed in Section II.B.3.b). As soon as they received approval from their institutional review boards and training from the Yale evaluation team, participating clinics began collecting patients' data using health surveys housed on the program's cloud-based data platform. Of the 13 clinics, 12 had begun collecting patients' data electronically by March 2015. One clinic did not collect data electronically because of institutional review board concerns about the online data platform. However, the institutional review board allowed this clinic to collect data through other means and share these data with the evaluation team. In May 2015, TCN received a no-cost extension of its award to support continued follow-up data collection at five participating clinics.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.3 summarizes the major facilitators of and barriers to TCN's implementation effectiveness in each domain.

Table II.A.3. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Ability to adapt the TCN program model to the unique needs of participating clinics • Perception of the program as an improvement upon standard care	• No significant barriers noted
Implementation process	• Successful engagement of community stakeholders	• Collection and use of patients' data for self-monitoring and quality improvement
Internal factors	• Team communication and collaboration • Clinics' previous experience serving the target population	• No significant barriers noted
External factors	• None	• Minimal resources available for targeted population

Sources: Interviews from second site visit, March 2015; document review, March 2015.

a. Program characteristics

Two characteristics of the TCN initiative helped program implementation: (1) the ability to adapt the TCN program model at participating clinics and (2) the perception of the program as an improvement upon standard care. First, as described in Section II.A.1.d, TCN's model can be adapted to suit the unique structures, staffing, and resources of participating clinics. For example, rather than require CHWs to follow specific outreach protocols to identify new patients, TCN helps each clinic develop a clinic-specific outreach plan using an outreach mapping tool. During the second site visit, one program leader noted that this approach "creates a structure that sites can use to identify the referral sources in their area but still gives them flexibility to figure out what makes the most sense for their [clinic]." Some clinics also hire a panel manager to collect patients' data in the online platform; other clinics use CHWs or other clinical or administrative staff based on their preferences and resources.

Second, staff on the existing primary care teams perceive the TCN program as useful and effective relative to their usual care practices in large part because of the added care provided by the CHWs. For example, during the second site visit, one TCN physician noted, "I couldn't live without CHWs … Often they go above and beyond. They are instrumental in making people feel comfortable with the health care system and helping people make it to their appointments." Another primary care physician at the same clinic commented that "It's invaluable for people to be able to come into the clinic within days of coming out of prison and get access to a primary care doctor and a mental health provider, and to be able to meet with the CHW who has been

where you've been and has made it and can show you that's it's possible." When asked what made the TCN program a success, a primary care physician at another clinic reported that it was a success "Just having someone paying attention to what's going on with people who are leaving the [correctional] system because it's usually a pretty poor transition for folks when they come out." Another physician stated that the program helps to address the needs of the patients because their needs are "very little medical and very much social."

b. Implementation process

One implementation process factor in particular facilitated the implementation of the TCN program: successful engagement of community stakeholders. As described previously, CHWs conduct outreach to community stakeholders to identify formerly incarcerated people who are in need of medical care and to connect patients with available services and resources for the target population. TCN program leaders and clinic staff noted that it is important to continue to maintain strong relationships with community stakeholders because these relationships are reciprocal; that is, these organizations are both a source of potential patients and a resource or support for existing patients. For example, CHWs can give a presentation on the TCN to a substance abuse treatment center with the hope of receiving new referrals, but also refer TCN patients to this organization when in need of drug treatment. In addition to social service and behavioral health treatment organizations, CHWs described the importance of building relationships with local parole and probation departments. The benefit of these stakeholder relationships was most visible in one of the newer clinics, which is housed within a county reentry resource center. The reentry center is intended to serve as a one-stop shop for people recently released from prison, with parole, probation, and many reentry support services located within the same building that houses the CHW and TCN clinic. This unique arrangement enables the CHW to both receive and make referrals to community partners. Staff at this clinic credit their easy access to and good relationship with community stakeholders as the primary reason for the clinic's success at meeting its enrollment goals.

One implementation process factor posed a challenge for implementation of the TCN program: collection of patients' data for those who consented to having their data entered into the online platform. As described previously, most clinics assigned data collection responsibilities to CHWs or other clinic staff, rather than creating a separate panel manager position. Although the evaluation team provided each clinic with training on how to collect patients' data using the surveys in the online data platform, TCN program leadership reported that clinic staff had trouble with data entry because staff were not sufficiently familiar with the online data platform or data collection in general. In addition, the follow-up data collection activities required significant time to complete; several staff noted that these activities were burdensome and difficult to complete in addition to providing needed services to patients. The evaluation team worked with clinics in several ways to address this challenge, such as (1) providing support by email, telephone, and the chat system in the online data platform to help answer questions about data collection or entry; (2) creating and regularly updating a frequently-asked-questions list that is available to clinic staff as a resource on the online data platform; and (3) holding additional panel manager trainings on specific topics, such as collecting follow-up data, data extraction, and specific survey tools the program uses to collect the data. TCN

program leadership also hosted a competition among the clinics to encourage staff to meet the follow-up goals, with gift certificates as a reward.

Despite these efforts, TCN clinic staff reported continued challenges engaging patients in data collection. For example, the participating clinics faced difficulty finding and engaging patients in follow-up data collection surveys for Yale University's evaluation of the program. The Yale evaluation team set a goal of collecting follow-up data from 70 percent of patients enrolled at each clinic who consented to having their data entered into the online platform. Over the course of the TCN program, the awardee's narrative progress reports described follow-up rates at individual clinics that ranged from 30 to 81 percent as of March 2015. TCN program leaders reported that, after they provided clinics with technical assistance, follow-up rates improved, although rates at some clinics were still below the 70 percent target.

c. Internal factors

Characteristics of the organization implementing a program can also affect implementation effectiveness. One internal factor in particular influenced implementation of the TCN program: collaboration and communication among primary care team staff in TCN clinics. The existing primary care teams in most clinics collaborated closely with CHWs to provide and coordinate patients' care. During the second site visit, primary care physicians noted that working with a CHW helps the physicians better understand patients and respond appropriately to their needs. For example, one physician noted that those recently released from a long stay in prison often are late to, or miss, medical appointments because they have trouble using public transportation and finding their way to the clinic. If a patient misses an initial appointment with a physician, the CHW can communicate an explanation to the physician, helping the physician to respond with empathy and ensuring that the patient receives the necessary care. Another physician said that the CHW helps to "bridge the gap" between the physician and his patients, "both culturally and in that her initial meeting with the patient isn't always in the clinic, so they have someone that they establish trust with from the start." He noted that patients often seem more at ease during their initial appointments after he mentions that he works with the CHW and commented that having the CHW on staff to meet with patients "has been really essential in establishing trust and in helping patients know that we are there to help them." One TCN program leader noted, "That's really the beauty of the CHWs. The work that you do becomes culturally competent on a patient-by-patient basis because you're working with the CHW." CHWs also recognize the dedication of other members of the care team. For example, one CHW described the TCN physicians as "passionate about their work and the care they give patients." CHWs at other clinics echoed this sentiment, with one CHW stating the physicians have "incredible dedication to working through the system and its barriers, strategizing together to find solutions, and creating a place where formerly incarcerated patients can see that they are being provided a unique service."

Another internal factor that might influence implementation is sites' prior experience implementing the program model or serving the target population. At least 7 of the 13 clinics were well prepared to serve the target population because of their previous experiences providing health care to patients recently released from prison and employing formerly incarcerated people. At one clinic in California, for example, CHWs with a history of incarceration have worked with

people returning from prison since the TCN program model was first implemented in 2006. Similarly, the second California clinic had prior experience with the TCN program model and, as of 2011, employed a CHW with a history of incarceration. After joining the TCN program, the two clinics began collecting more targeted data on the patients and focusing more on improving the quality of care. Although several other clinics had prior experience serving people who were returning from prison, they did not have experience implementing the TCN program model. For example, one clinic began providing such services—and employed social workers who were formerly incarcerated—in 2010.

d. External factors

Features of the environmental context in which the organization is located can also influence implementation effectiveness. One external factor, the minimal resources available for the targeted patient population, posed an implementation barrier. Many people return to society after long periods of incarceration to find that the health care system—along with many other things, such as transportation and technology—has changed dramatically. During the second site visit, clinic staff lamented the complexity of the corrections and health care systems, noting "I have a master's degree and I don't understand half of the forms they bring in. It's frustrating and upsetting and [patients] need to have so much patience and courage to make it through." Patients also face barriers in other areas of their lives that relate to their ability to engage in health care services. For example, after release from prison, many of them do not have a place to live and are placed in halfway houses or homeless shelters in neighborhoods where drugs and crime are rampant and very few resources are available. One physician commented, "They're expected to live on food and shelter alone, and they have no way to visit their families." Nearly all clinic staff cited homelessness as one of the biggest challenges preventing patients from continuously engaging in care.

Although these factors pose barriers to successfully engaging and retaining patients in the TCN program, clinic staff understand the potential hurdles and arrange clinic resources and services in such a way to meet their needs. First, clinic staff create a welcoming environment for patients. As one physician noted during the second site visit, "the world out there can really feel like a scary place, and coming here might feel like the only place [patients] can feel safe. They can talk to someone who isn't going to yell at them or judge them. It's a moment they have to really breathe and be able to deal with the rest of their lives." Second, CHWs provide care coordination and navigation services, such as accompanying patients to other health and social service appointments and assisting them in applying for housing. For example, when describing the benefit of the clinic to patients she reaches out to in homeless shelters whom she assessed as being "completely stressed out," one CHW commented, "Having the clinic helps to make sure they don't have to worry about figuring out when their next appointment [is] or how to get their medication if they run out before that appointment. They know that they can contact me and I can email the doctor and help figure it out." Third, CHWs also work with other clinic staff to identify new ways to meet these challenges at the clinic level. For example, several clinics are now offering or planning to offer patient support groups, after one CHW at a participating clinic successfully initiated such a group that meets weekly and regularly attracts a core group of patients. The CHW noted that the support group is "a place for people who need to talk about the

prison experience and about their day-to-day experiences. [Patients] really need aftercare treatment coming back into the community, just like if you're coming out of inpatient [medical] treatment."

4. Sustainability and scalability

All of the participating clinics plan to continue using and adapting the TCN program model to serve formerly incarcerated patients after HCIA funding ends. To do so, clinics must identify funding sources to cover the employment of CHWs and maintain buy-in from participating clinical staff and institutional leadership (for example, educating new organizational leadership or staff about the program to ensure that it remains a priority). TCN program leaders reported that most clinics have identified funding sources to support ongoing implementation or continuation of the CHW position. However, as of March 2015, two clinics were still in the process of identifying potential funding sources. Some clinics plan to continue staffing the CHW as part of the primary care team, but with modified responsibilities. For example, as a result of changes within the broader system in which it is housed, one clinic plans to keep the CHW on staff but might move this person into a general case manager position that will serve other patients in addition to those recently released from prison.

The TCN program administrators plan to sustain several other supporting activities beyond the award period, such as the online data platform, annual training retreats, and some of the online courses developed using HCIA funding. TCN is also using funding from the Langeloth Foundation to convene its National Advisory Board to discuss issues related to program sustainability, such as (1) the professionalization of the CHW role and barriers to hiring formerly incarcerated staff to serve as CHWs; (2) using Medicaid funding to support TCN clinics and CHWs in response to a Centers for Medicare & Medicaid Services (CMS) final rule on Medicaid and the Children's Health Insurance Program that, as of January 2014, allows state Medicaid agencies the option to provide reimbursement for preventative health services provided by CHWs (U.S. Congress 2013), and (3) sustaining and expanding the post-prison health worker certification program.

The program model funded through the HCIA is an expansion of the TCN program model started in 2006 at a clinic in California and is currently operating in the 13 clinics participating in the TCN program, as well as 3 additional sites in California and one in Arkansas that began implementing the program model with other funding in the third year of the award period. Throughout the award period, the program partners provided technical assistance to other clinics interested in adopting the TCN program model outside of the HCIA-funded program. TCN program partners plan to continue to expand to new clinics and pursue several activities to support scaling the TCN program model. These include technical assistance and training (for example, online cultural competency training) and ongoing monthly meetings with TCN liaisons at new and existing clinics to facilitate mentorship and sharing of lessons learned. TCN program leaders noted that several clinics not currently participating in the TCN program have expressed interest in the program. Program leaders are in the process of writing grants to obtain the funding needed to expand the TCN model to these additional clinics and provide related technical assistance.

In addition, TCN used HCIA funding to develop several tools and products that interested health centers or hospitals can use to support implementing the program model. These include (1) a needs assessment tool to assess potential clinics for readiness to implement the TCN program model, (2) CHW job descriptions and other documents to assist clinics with hiring formerly incarcerated staff, (3) an outreach mapping tool and outreach plan to assist CHWs in identifying community stakeholders for outreach and referral, and (4) an assessment tool to facilitate CHWs' integration and supervision.

The TCN program partners also are leveraging other elements of the training program that were expanded using HCIA funding to support continued training of CHWs in transitions clinics and other settings. First, CCSF is creating a facilitators' guide to supplement its existing CHW textbook, *Foundations for Community Health Workers*, and updating the textbook with 5 new chapters. The 25-chapter online facilitators' guide will contain more than 150 training activities and assessment resources (such as quizzes, case studies, and grading rubrics) tied to each chapter of the textbook; it will also integrate more than 90 online video resources developed using HCIA funding. From January to March 2014, CCSF solicited feedback from external CHW training experts on a draft version of the facilitators' guide and, in November 2014, the publisher issued a pre-release of 5 chapters of the facilitators' guide for additional feedback from CHW trainers who use it. During the second site visit, one TCN program leader commented, "Road-testing [the facilitators' guide] helps us understand whether it's going to be something people will use and if it will meet their needs." During the rest of the award period, CCSF plans to continue to work with the publisher to finalize the online facilitator guide and related textbook content.

B. Clinicians' Attitudes And Behaviors

1. HCIA Primary Care Redisign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected clinics provided important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff are limited to a small number of clinic locations and might not reflect the perspectives of clinicians practicing at other sites. In order to assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians at the participating clinics in fall 2014, the third year of the HCIA-funded program. Data from the survey provide additional insights into the implementation process and experience as well as the contextual factors that might affect implementation effectiveness at the 13 clinics implementing the TCN program.

In this section, we report on the views of clinicians of their daily work lives and practices. First, we focus on the contextual factors that can affect program implementation, including the characteristics of the practice locations, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care teams function. We then present data on the alignment of TCN clinicians' views and experiences with the overall goals of the HCIA-funded innovation, as well as their awareness of and participation in the TCN program and their view of the facilitators of and barriers to successful program implementation. Throughout this section, our findings exclude survey and item

nonrespondents, as well as clinicians who reported that a given question did not apply to their practices and thus did not provide responses.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

A total of 17 TCN clinicians responded to the survey, resulting in a response rate of 78 percent. Given the small sample size, we describe the results but do not present raw data. Of the respondents, nearly all were physicians. More than half of the respondents reported that they practice at community health centers (including FQHCs); of the remaining respondents, most characterized their practice locations as a medical school or a private-/not-for-profit-run university or hospital. More than three-quarters of clinicians reported earning a fixed salary, and slightly more than 10 percent of clinicians earned salaries adjusted for performance.

TCN clinicians reported working in settings that are advanced in terms of health IT. Although nationally, slightly more than half of clinicians practice in settings with functional EHRs, all TCN clinicians reported using health IT at their practice locations (Furukawa et al. 2014). This finding is not surprising, given that TCN selected clinics based in part on their data collection capacity. Nearly all clinicians reported using electronic systems for entering clinical notes, receiving drug dosing or interaction alerts, and accessing laboratory test results; more than three-quarters reported that they used electronic systems to prescribe medications and order tests and procedures. In addition, most clinicians reported using electronic referral tracking systems and patient registries—functions that are generally advanced and not in widespread use nationally (DesRoches et al. 2014). However, relatively few TCN clinicians reported that they offer patient-facing technologies. Fewer than half of respondents reported offering their patients the option to email a clinician about a medical question or request prescription refills or appointments online.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall careers, level of burnout, and perceptions of their practice environment can all have an effect on the success of program implementation and organizational change. Most TCN clinicians who responded to this survey reported that they were at least somewhat satisfied with their careers in medicine; more than a quarter of respondents said that they were very satisfied. None of the respondents reported experiencing persistent or overwhelming burnout; however, most clinicians said that they experienced occasional stress and slightly more than a quarter reported they experienced some symptoms of burnout at the time the survey was taken.

TCN clinicians gave similar ratings to their workplace management (Table II.B.1). More than three-quarters of respondents agreed that their management teams were supportive, they were encouraged to offer suggestions and improvement, and they had adequate opportunities for professional development. TCN clinicians had a mixed response to the amount of work they were expected to complete each day. Although most respondents agreed that the expectations were reasonable, more than a third disagreed with this statement.

Table II.B.1. Workplace ratings

A majority of clinicians strongly agreed that:
• I have adequate opportunities to develop my professional skills.
A majority of clinicians at least somewhat agreed that:
• Management is supportive of me. • I feel encouraged by my supervisor to offer suggestions and improvements. • The amount of work I'm expected to finish is reasonable. • It is possible to provide high quality care to all of my patients.

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: A total of 17 clinicians responded to the survey. In each answer category, there were fewer than 11 responses. Hence, raw data have been withheld because of confidentiality restricitions.

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. Most responding clinicians agreed with the statement "It is possible to provide high quality care to all of my patients." Most TCN clinicians reported that major barriers to providing optimal care were difficulties obtaining specialist referrals for patients in a timely manner, patients' inability to pay for care, insufficient reimbursement, lack of timely information about care provided to patients by other clinicians, and lack of time to spend with patients.

c. Clinicians' perceptions of care team functioning

All TCN clinicians who responded to this survey reported working as part of a care team; overall, their perceptions of how these teams functioned were positive. More than three-quarters of TCN clinicians agreed that care team members relayed information in a timely manner, had sufficient time for patients to ask questions, used common terminology when communicating with one another, and verbally verified information they received from one another. Slightly fewer respondents—although still a majority—agreed that care team members followed a standardized method of sharing information when handing off patients.

d. Alignment with goals of PCR

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all (Table II.B.2). The inclusion of the extremely important category forces respondents to choose between goals that are essential to meet and those that are merely important, thereby identifying the goals respondents feel are most important for PCR. Based on the proportion of clinicians rating each of these goals as extremely important, the views of TCN clinicians appear to generally align with the goals of PCR. Most clinicians rated 8 of the 13 goals as extremely important; the 2 goals that were endorsed by the largest majority of clinicians were increasing access to primary care and improving care coordination for patients with chronic conditions, which are perhaps the two most central goals of the TCN program. Slightly fewer than half of clinicians identified the remaining five goals—such as reducing hospital readmissions, reducing ED visits, and reducing overall health care spending—as extremely important.

Table II.B.2. Importance of PCR goals

Goals rated by most clinicians as extremely important	Goals rated by many clinicians as extremely important	Goals rated by few clinicians as extremely important
• Improving the capability of health care organizations to provide patient-centered care • Improving capability of health care organizations to provide team-based care • Improving appropriateness of care • Increasing access to primary care • Improving care coordination for patients with chronic conditions • Improving patients' capacity to manage their own care • Increasing the use of evidence-based practices in clinical care • Improving care continuity in primary care	• Reducing hospital readmissions • Reducing ED visits • Reducing overall health care spending • Increasing the number of primary care practices functioning as a PCMH • Increasing use of electronic health records and other health IT	• None (all goals rated as extremely important by at least one-third of respondents)

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: A total of 17 clinicians responded to the survey. In each answer category, there were fewer than 11 responses. Hence, raw data have been withheld because of confidentiality restrictions. *Most* refers to statements endorsed by the majority of respondents; *many* refers to statements endorsed by more than one-third of respondents; and *few* refers to statements endorsed by fewer than a third of respondents.

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the TCN program is to change the way care is provided to people transitioning from prisons to communities. Program administrators believe that clinicians are critical to that process. Understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on patients' outcomes. For example, if clinicians are aware of the TCN program, have received appropriate and effective training, and believe that the program will have a positive effect on the care they provide, they are likely to feel more invested in the program's success. Alternatively, those who feel more negatively about the TCN program might be less likely to enthusiastically implement it. In this section, we report on clinicians' experiences with and perceptions of the TCN program.

a. Awareness of the program and receipt of training

Nearly all of the surveyed clinicians were at least somewhat familiar with the TCN program. Of these clinicians, about three-quarters had received training related to the program. Clinicians reported that they received 2 to 40 hours of program-related training, with an average of 19.6 hours of training.

b. Perceived effect of program on patients' care

Most clinicians reported that they believe the TCN program positively affected patients' care across all of the dimensions surveyed. Clinicians were asked about the perceived effect of the TCN program and the barriers to and facilitators of implementation only if they reported being at least somewhat familiar with the program. More than three-quarters of clinicians who

were familiar with the TCN program believed the program positively affected the quality and patient-centeredness of the care they provide, as well as the equity of care for all patients and the ability to respond to their needs in a timely way. A slightly smaller majority of clinicians familiar with the TCN program believed the program positively affected efficiency and safety of patients' care. None of the clinicians who responded to this survey perceived negative impacts of the program; rather, a small number believed the intervention would have no effect on the care they provide or that it was simply too soon to tell if there would be effects.

c. Barriers and facilitators to program implementation

Finally, clinicians who were at least somewhat familiar with the TCN program rated the effect of a series of barriers to and facilitators of program implementation (Table II.B.3). Respondents most frequently cited as a facilitator of program implementation the quality of interpersonal communications with allied health professionals, such as CHWs. This finding reflects in-person discussions with clinicians, who emphasized the integral nature of CHWs to the TCN program. Most respondents also reported that the availability of personnel and relevant patient information at the point of care positively affected program implementation, as did the required use of computer and communications technology and the availability of community resources for patients with complex conditions. Few clinicians identified any barriers to implementation; the amount of required documentation was most often cited by clinicians (though reported by fewer than one-fifth of responding clinicians) as a barrier to program implementation.

Table II.B.3. Facilitators to program implementation

Factors rated by most clinicians as facilitators:	Factors rated by many clinicians as facilitators:	Factors rated by few clinicians as facilitators:
• Availability of personnel • Availability of relevant patient information at the point of care • Required use of computer and communications technology • Availability of evidence-based clinical information • Availability of community resources to care for patients with complex conditions • Quality of interpersonal communications with other allied health professionals	• Level of program funding • The amount of time required by the program • Quality of interpersonal communications with other providers	• Amount of required documentation • Quality of interpersonal communications with specialists

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: A total of 17 clinicians responded to the survey. In each answer category, there were fewer than 11 responses. Hence, raw data have been withheld because of confidentiality restrictions. *Most* refers to statements endorsed by the majority of respondents; *many* refers to statements endorsed by more than one-third of respondents; and *few* refers to statements endorsed by fewer than one-third of respondents.

4. Conclusions about clinicians' attitudes and behavior

Surveyed clinicians generally responded positively to the TCN program model. These findings align with clinicians' experiences at the four clinics visited during two rounds of site visits. Most clinicians reported that the TCN model enabled them to provide high quality care to patients. The majority of clinician respondents also reported that the care teams (including other clinicians and CHWs) worked well together across all dimensions of care and that they had good communication with CHWs. Few surveyed clinicians reported that they found the data collection for the program to be burdensome. Finally, the survey findings suggest that clinicians believe in the TCN mission. Clinicians' perceptions of the most important primary care redesign goals—improving primary care access and care coordination—reflect the goals of the TCN program. Both in survey responses and interviews, TCN clinicians reported that they believe the program will help them to provide more patient-centered and equitable care to indiviuals recently released from prison.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

TCN received HCIA funding to expand its existing program model to additional clinics around the United States. TCN aimed to improve health and health care and lower costs for people with chronic health conditions who were recently released from prison by providing and coordinating primary care and other services. To support the program, TCN program administrators developed an online version of an existing CHW training program for people with a history of incarceration. They also implemented an online data platform through which clinics could collect and use patient data. After nearly three years of HCIA funding, TCN program administrators and clinics have largely succeeded in implementing the program model and activities to support it. Program implementation was facilitated by clinics' ability to adapt the model to best suit their unique resources and staffing structures, close collaboration among TCN care team members, and CHWs' ability to create and leverage connections with external community stakeholders. Challenges facing people recently released from prison, such as housing and transportation, which often kept them from getting to the clinics to receive services, hindered implementation. Clinic staff also had difficulty using the online data platform effectively, despite the training they received from the local evaluation team. The HCIA Primary Care Redesign Clinician Survey found that most clinicians at participating clinics believed the TCN care teams functioned well and that the TCN program positively affected patient care quality.

Our next steps for this evaluation are to (1) monitor TCN's program implementation reports through June 30, 2016, and its plans for sustaining the program beyond the funding period; and (2) evaluate trainees' attitudes and experiences with the program in the third year of the award through an administered survey. We were unable to conduct an impact analysis for this awardee because we could not construct a valid comparison group for TCN's unique and vulnerable treatment population, nor could we measure program outcomes for the treatment group before program implementation.

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

REFERENCES

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no. 9, 2014, pp. 1672–1679.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

U.S. Congress. "Final Rule: Medicaid and Children's Health Insurance Program: Essential Health Benefits in Alternative Benefit Plans, Eligibility Notices, Fair Hearing and Appeal Processes, and Premiums and Cost Sharing; Exchanges: Eligibility and Enrollment." Section 42 CFR §440.130. 78. *Federal Register,* vol. 78, no. 135, July 15, 2013. Available at http://www.gpo.gov/fdsys/pkg/FR-2013-07-15/pdf/2013-16271.pdf. Accessed August 2, 2015.

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

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Pacific Business Group on Health

March 2016

Rosalind Keith
Rumin Sarwar
Boyd Gilman
Catherine DesRoches
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850
Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Lorenzo Moreno

Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

PACIFIC BUSINESS GROUP ON HEALTH

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by the Pacific Business Group on Health (PBGH) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the PBGH program and Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, in Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF PBGH

PBGH, a nonprofit business coalition, received a three-year, $19.1 million award to implement the Intensive Outpatient Care Program (IOCP) beginning in August 2012. PBGH provided technical assistance to 23 participating medical groups (PMGs) in five states to implement the IOCP. Table I.1 summarizes key features of the program. Through the IOCP, PBGH aimed to reduce hospitalizations and emergency department (ED) visits and lower the total cost of care by 5 percent each; it also intended to improve program participants' experiences, health care quality, and health status. HCIA funding for the IOCP ended in June 2015, although PBGH received a partial no-cost extension to provide technical assistance for sustaining an IOCP model for Medicaid beneficiaries.

Table I.1. Summary of PBGH PCR program

Awardee's name	Pacific Business Group on Health
Award amount	$19,139,861
Implementation date	August 2012
Award end date[a]	June 2015
Program description	Strengthen PMGs' capabilities for identifying medically complex patients and providing them with personalized care management services
Innovation components	Care management
Intervention focus	Patient
Workforce development	Create new positions and change roles and responsibilities of existing staff to embed care management services in primary care practices
Target population	Medicare beneficiaries with chronic conditions and/or frequent utilization
Program setting	Provider-based (primary care practices)
Market area	Multistate (Arizona, California, Idaho, Nevada, and Washington)
Market location	Various
Core outcomes	• 5 percent reduction in hospitalizations • 5 percent reduction in ED visits • 5 percent reduction in total cost of care • 2 to 4 percent improvement in participants' experiences • 2 percent improvement in health care quality and health status, as measured by condition-specific indicators

Source: Review of PBGH program reports, March 2015.

Note: The implementation date represents when PMGs began taking steps toward launching the program by hiring and training staff and undertaking other operational activities related to program implementation.

[a] HCIA funding for the IOCP ended in June 2015, although PBGH received a partial no-cost extension to provide technical assistance for sustaining an IOCP model for Medicaid beneficiaries.

II. SUMMARY OF FINDINGS

In this chapter, we summarize the methodology and present the main findings of the evaluation as they relate to (1) program implementation, (2) clinicians' attitudes and behaviors, and (3) participants' outcomes.

A. Program implementation

In this section, we provide a detailed description of the IOCP, highlighting changes in program design over time. Second, we review the evidence on implementation effectiveness, including an assessment of measures of enrollment, the IOCP implementation timeline, and other service- and staff-related metrics used by PBGH to monitor implementation. Third, we examine the facilitators and barriers that can influence implementation effectiveness, specifically those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss PBGH's plans for program sustainability. We based our evaluation of PBGH's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with the awardee, and information collected during site visits conducted in April 2014 and April 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

The IOCP had a single component: providing care management services to high-risk patients. In this report, the term *care managers* refers to licensed health professionals staffed on IOCP care teams, although implementing sites used various titles to refer to IOCP care team members. After participants enrolled into the program, care managers interacted with them one on one to learn about their medical and social needs, provide them and their caregivers with education and emotional support, and connect them to appropriate community resources. Knowing that a one-size-fits-all implementation strategy would not be appropriate given the diversity of PMGs, PBGH granted PMGs the freedom to adapt the program to their specific needs while adhering to the programs' guardrails. Over the course of the award, PBGH did not change the original design of the IOCP model.

b. Target populations and participant identification, recruitment, and enrollment

Table II.A.1 provides key details about the target populations and the participant identification, recruitment, and enrollment processes for the IOCP. Participants targeted by PMGs to receive IOCP services included chronically ill Medicare beneficiaries (including those dually eligible for Medicare and Medicaid) who were at high risk of experiencing a hospitalization. PBGH did not change the target population since the program launched, but it changed the enrollment target from 27,000 to 15,000 in the ninth quarter (July through September 2014) because of challenges in meeting the original enrollment target.

Table II.A.1. Key details about program design and adaption

	Care management
Target population	Medicare beneficiaries (including dually eligible beneficiaries) with chronic illnesses who are at high risk of experiencing a hospitalization
Identification strategy	Direct referrals by primary care physicians, transfers of participants from existing care management programs, the use of patient data to identify high-risk participants, and other PMG-developed strategies
Recruitment/ enrollment strategy	Participant enrollment (intent to participate) can take place either via telephone or letter. The participant must sign a consent form. To recruit, PMGs use the following approaches: (1) "warm hand-off" approach, in which primary care providers (PCPs) introduce high-risk patients to the care manager; (2) cold-calling patients from risk-stratification lists; and (3) other strategies, such as approaching high-risk patients during their stay in a hospital or skilled nursing facility.
Service delivery protocols	Within the first month of enrollment, each participant must complete three assessments that are used to create the personalized care plan, also known as the shared action plan. Participants are enrolled for a minimum of 12 months. Care managers typically meet with participants in person (in the participant's home or other location) at least once. Care managers must contact participants at least once a month—by telephone, in person, or online—for the duration of a participant's enrollment.
Adaptations	Yes; PBGH originally expected PMGs to identify high-risk participants by using risk scores calculated by the Milliman Advanced Risk Adjustors model. The varying quality of data submitted by the PMGs to Milliman resulted in the risk scores being delayed and, therefore, PMGs developed internal strategies for identifying high-risk participants. PBGH also adapted the enrollment strategy to have PCPs recruit participants in person after the initial cold-calling approach was not well received by patients.

Sources: Interviews from second site visit, April 2015; document review, March 2015.

PBGH did adapt the patient identification and enrollment processes. Originally, PBGH expected all PMGs to identify high-risk patients using the Milliman Advanced Risk Adjustors model to calculate risk scores based on PMGs' Medicare fee-for-service claims. (Milliman is a consulting and actuarial firm that processes Medicare claims through a proprietary algorithm to provide health care providers with risk scores for their patients.) However, the complexity of the claims-reporting requirements and the varied quality of the claims data submitted by PMGs delayed the development of risk-stratified patient lists. Therefore, PBGH developed alternative methods for identifying high-risk patients, including the following:

1. Direct referral by a primary care physician
2. Transfer of partients from existing care management programs
3. Identification through hospital records of patients with three or more hospitalizations or ED visits in the past six months
4. Identification through internal reporting of patients who saw three or more specialists, had three or more active monitored conditions, or patients who were on five or more medications
5. Other ways that supported successful enrollment, such as identification of high-risk patients during their stay in a hospital or skilled nursing facility

In addition, PBGH developed a set of protocols to improve the quality of the data submitted to Milliman by the PMGs. These included (1) organizing in-person working sessions with the data teams at each PMG (and follow-up sessions in person and via teleconference); (2) holding conference calls with each PMG and Milliman to ensure each PMG understood data submission specifications and data extract, transform, load (ETL) processes; (3) distributing a thorough data training manual; and (4) performing a preliminary quality audit before data submission. However, by the time PMGs resolved their claims data issues and Milliman began preparing their risk scores, PMGs had successfully implemented the alternative methods of participant identification. As a result, most PMGs did not incorporate the Milliman risk scores into their workflows.

PMGs also experienced challenges with participant enrollment, which required additional program adaptations. Originally, members of the care management teams reached out by cold-calling patients to introduce the program. Patients were not receptive to this approach due to unfamiliarity with the care management team members, so PMGs implemented new strategies to enroll them. Patients were most receptive to the warm hand-off approach, in which PCPs introduced high-risk patients to the care manager. In addition, after encountering difficulties enrolling patients over the telephone, one site began enrolling patients during their hospital stays, which enabled care managers to meet patients in person. Meeting patients at the hospital also reduced the time and burden care managers faced when traveling to multiple participants' homes.

c. Service delivery protocols

PBGH developed program requirements (called guardrails) for the PMGs to use as a guide during program implementation. The most intensive contact with participants occurred in the first month of enrollment, when participants must complete three standardized assessments to remain enrolled in the program. The results of the assessments helped the care managers (usually a registered nurse, social worker, and/or clinical pharmacist) develop a personalized care plan, also known as a shared action plan, for the participant. Care managers contacted patients at least once a month—by telephone, in person, and occasionally through online communication. During these encounters, care managers and participants jointly updated the participant's shared action plan as the participant's health progressed. Additional guardrails are listed in Table II.A.2.

Program staff typically found that participants' conditions stabilized after the first few months of enrollment, after which contact with participants dropped to every other week or once a month. Care managers disenrolled participants from the IOCP for any of the following three reasons:

1. **Graduation.** The participant completed 12 months of enrollment in the IOCP and successfully completed the shared action plan or stabilized his or her condition.

2. **Drop out.** The participant died, declined to continue participation, or was lost to follow-up.

3. **Enrolled in error.** Care managers discovered the participant was ineligible after enrollment.

Table II.A.2. Additional PBGH guardrails for the IOCP implementation

Guardrail	Description
Care management model	PMGs can choose to implement a distributed model, an intensivist model, or a hybrid of the two models. The distributed model involves embedded care managers working with multiple primary care physicians (PCPs) to maintain a specific IOCP caseload. The intensivist model involves a PCP practice dedicated to care management, meaning IOCP participants might be required to change their PCP to a PCP in that practice. In the hybrid model, PMGs have some practices that incorporate embedded care managers and some that are dedicated to care management.
Care management program staff	IOCP care management program staff cannot be assigned to other clinical functions that would conflict with IOCP.
IOCP care teams	IOCP care teams consist of one PCP and three care managers, two of whom must be registered nurses. Care team members must receive culturally and linguistically appropriate training as needed.
PCP—participant attribution	Each participant must have a PCP and the PMGs must enlist PCP support for the program and actively involve them in participant enrollment and care management.
Determination of risk distribution cut-point for patient list	Risk distribution cut-points for appropriate patients will be decided by the PMGs in conjunction with PBGH using Milliman scores.
Patient outreach	Patient outreach will ideally be conducted by someone the patient knows and who can make a warm hand-off of the patient to care manager.
Participant intake	Participant must sign a consent form to participate in the IOCP. Participant intake must occur during a one-on-one, face-to-face so-called super visit, within a month of the participant enrollment, and ideally within one or two weeks.
Required participant assessments	Three participant assessments must be administered within one month of enrollment: (1) Patient Health Questionnaire-2 (PHQ-2), Patient Activation Measure (PAM), and Veterans RAND 12 Item Health Survey (VR-12).
Additional participant assessments	Additional assessments include depression, pain, back pain, and advance directive. Follow established clinical protocols for delivery.
Participant contact	Contact (defined as two-way interaction) between the care manager and participant must occur at least once per month by telephone or in person. The care manager must develop a relationship with the participant as early as possible.
Participants' goals	Every participant must have at least one goal per year. Goal(s) should be specific, measurable, and attainable.
Participant access	Participants must have access to a care manager (or other staff) who has access to the participants' medical records (access can be fulfilled in different ways, as long as same-day notification can be made to the care manager).
Secure messaging	Care managers must have access to secure messaging during working hours.
Support services	The care team must develop a comprehensive list and relationships with support services from which the participant will benefit, such as home health and durable medical equipment, behavioral health, substance abuse, community care, Meals on Wheels, senior centers, area agencies on aging, and others as needed. Actively plan to address transportation needs, which can represent a significant barrier to participants achieving their goal(s), access to care, and/or self-management.

Table II.A.2 *(continued)*

Guardrail	Description
Compliance with contractual requirements	The PMG must comply with PBGH contractual requirements, including measures requirements and CMS contractual requirements.

Source: Interviews from second site visit, April 2015; document review, March 2015.

d. Intervention staff and workforce development

PBGH uses its award to support some staff positions. Approximately 38 percent of the award funding is allocated to PBGH and PMG to support personnel to design and build the program, and PMGs receive funding to staff an information technology analyst and a project manager. Approximately 27 percent is allocated to subcontractors to conduct workforce training, build a care management platform, manage data systems and reporting, and conduct evaluation. Approximately 25 percent is allocated to PMGs to subsidize care management and data reporting.

IOCP staff included a combination of licensed and non-licensed personnel, including registered nurses (RNs), social workers, clinical pharmacists, medical assistants, and licensed practical nurses (LPNs) (Table II.A.3). Licensed care managers (RNs, social workers, and clinical pharmacists) conducted the initial consultations with enrolled participants and led care teams in assessing participants' needs and developing shared action plans. Nonlicensed program staff (medical assistants) were important members of the care team as well, especially because of their ability to connect participants to needed community resources. When developed, the nonlicensed program staff often took the lead in executing the shared action plan, by connecting participants to community resources, regularly following-up with participants, and monitoring participants' conditions.

Table II.A.3. Key details about program staff

Staff members	Staff/team responsibilities	Adaptations?
Registered nurses (RN)	Conduct initial consultation with enrolled participants; lead care teams in assessing participants' medical needs and developing shared action plans	No
Clinical pharmacists	Conduct initial consultation with enrolled participants; lead care teams in assessing participants' prescription drug needs and developing shared action plans	No
Social workers	Conduct initial consultation with enrolled participants; lead care teams in assessing participants' psychosocial needs and developing shared action plans	No
Medical assistants	Execute the shared action plan by connecting participants to community resources, regularly follow-up with participants, and monitor participants' conditions	No
Licensed practical nurses (LPN)	Execute the shared action plans by connecting participants to community resources, regularly follow-up with participants, and monitor participants' conditions	No

Sources: Interviews from second site visit, April 2015; document review, March 2015.

PBGH employed a variety of strategies to provide technical assistance to PMGs with IOCP implementation and to train program staff (Table II.A.4). PBGH maintained close relationships with many PMGs via quarterly leadership trainings. During these training sessions, PMG administrators discussed IOCP implementation experiences and shared best practices, particularly for overcoming challenges such as participant enrollment. PBGH required all frontline staff delivering direct care management services to attend a care management academy training. The academy was designed to teach frontline staff about program requirements and basic principles of care management, including participants' psychosocial issues, motivational interviewing, goal setting, participant assessments, and participant engagement. Based on feedback from PMGs, PBGH expanded training for program staff in the sixth quarter (October through December 2014) to include peer clinical case conferences. At these conferences, newly hired care managers attended training and existing care managers discussed challenging clinical cases and shared success stories with their peers under the facilitation of IOCP clinical advisors.

Table II.A.4. PBGH technical assistance and training activities

Technical assistance	
California quality collaborative leadership meetings	PMG administrators attended these quarterly meetings, during which they shared best practices for implementing the program.
Process improvement workshops	PBGH required all PMGs to participate in on-site workshops to identify specific actions and detail processes that facilitated participant enrollment and improved fidelity to the care management model through adherence to IOCP guardrails.
Workforce development	
Care manager academy	PBGH required all IOCP direct service staff to attend this training academy, at which they learned about program requirements (guardrails), participants' psychosocial issues, motivational interviewing, goal setting, and participant assessments and engagement. All IOCP direct service staff hired before July 1, 2014, attended the academy. The senior manager of clinical redesign at PBGH worked with all IOCP direct service staff hired after July 1, 2014, to provide the training.
Care coordinator office hour webinars	Webinars complemented the care manager academy, which covered topics such as end-of-life, burn-out, and motivational interviewing. Office hours were topic-driven, with both didactic and interactive components and were regularly attended by 50 to 100 people.

Source: Review of PBGH program reports, March 2015.

2. Implementation effectiveness

In this section, we examine the evidence on IOCP implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness of IOCP implementation (Table II.A.5). To assess effectiveness, we rely on interviews with program administrators, review of self-reported information in PBGH's quarterly self-monitoring and measurement reports, and site visits conducted in April 2014 and 2015.

Table II.A.5. PBGH self-reported program implementation measures

Measure	Target	Actual	Met target?	Adaptation?
Program enrollment	15,000	Approximately 15,008 (as of June 20, 2015)	Yes	Yes, changed original enrollment target, identified new participant identification strategies, and recruited participants by asking providers to introduce participants to the program (instead of care managers cold-calling patients)
Total participant encounters	Not specified	93,092 (as of December 31, 2014)	--	Some PMGs reduced the number of home and face-to-face visits to manage their caseloads
Average number of encounters per participant	Not specified	12.8 (as of December 31, 2014)	--	No
Percentage of participants with shared action plan	100%	92% (as of March 31, 2015)	No	No
Caseload (participants per care manager)	125	20–60	No	No; PMGs believe the target caseload is too high for the target populations
Program staffing	211.00 FTE by June 2015	214.25 FTE by December 2014	Yes	No, although two sites hired additional staff to increase outreach and recruitment

Sources: Interviews from second site visit, April 2015; document review, March 2015.

Note: We do not have data to understand why caseloads were so far below the targets; although we only interviewed a subset of PMGs and primary care practices on our site visits, those interviewed on our second site visit suggested they had flexibility to set their own staff and caseloads to meet organizational and patient needs.

FTE = full-time equivalent.

a. Program enrollment

By the end of the program (June 30, 2015), PBGH reported cumulative program enrollment of 15,008 participants, slightly above its enrollment target of 15,000 participants. The most recently reported quarter, the 12th quarter (April through June 2015), represents the smallest increase in enrollment (16 percent) from quarter to quarter during the award. Although the IOCP ended in June 2015, many PMGs continue to enroll participants and offer similar care management services to participants.

b. Program time line

PBGH tracked many of its implementation activities and whether they met the established timeline. Examples of implementation activities that met the established timeline include enrolling participants by the program launch date (May 1, 2013); holding multiple care coordinator academies and leadership training events by the end of the first year of

implementation; and successfully recruiting PMGs to participate in the IOCP program by the first quarter.

PBGH faced one exception in meeting the established timeline for implementation. PBGH originally planned for PMGs to identify patients using the Milliman Advanced Risk Adjustors model to calculate risk scores based on Medicare fee-for-service claims. As previously described, the complexity of Milliman claims-reporting requirements and the varied quality of the claims data submitted to Milliman by PMGs resulted in delays in developing risk-stratified patient lists. The first eight PMGs to receive Milliman reports received them in the seventh quarter (January through March 2014), three quarters after they began enrolling (April to June 2013). When available, PMGs found the Milliman risk-stratified patient lists unhelpful because the data used to create the lists were three to six months old. Instead, PMGs developed their own mechanisms to identify participants that better suited their participant populations and participating physicians.

c. Service measures

By the end of December 2014, IOCP staff had a cumulative total of 93,092 encounters with program participants, including telephone calls, in-person visits, and online communication (Table II.A.6). PBGH first reported the percentage of encounters that occurred in person in the fourth quarter (April to June 2013) when almost 42 percent of encounters occurred in person. Starting with the seventh quarter (January through March 2014), in-person encounters dipped and remained below 20 percent, although the average number of in-person encounters per person was constant across all quarters (less than one per month). For telephonic encounters, however, the average number per participant per quarter increased from the fourth quarter (less than one) to the seventh quarter (more than two). The upward trend is expected because new participants receive an in-person visit but are contacted mainly by telephone during the remainder of their enrollment. Also starting with the seventh quarter, the quarterly total number of encounters was four to six times higher than previous quarters, and the number of encounters per participant also increased as PMGs improved their efficiency, often by hiring more care managers, providing more training to care managers, or decreasing the indirect-service responsibilities of care managers. On average, a participant received 12.8 encounters with his or her care manager over the course of enrollment in the IOCP (at least 12 months). Care managers infrequently relied on electronic communication or audio/video conference to contact participants, a quarterly rate that fluctuated between 0.6 and 6.0 percent.

PBGH provided PMGs access to various performance data measures on a quarterly basis to help PMGs identify their strengths and troubleshoot weaknesses. For example, a PMG with low enrollment rates would know to change its enrollment approach and borrow strategies from PMGs that met their enrollment targets. Similarly, knowing the cause of disenrollment helped some PMGs recraft how they introduce patients to the program to improve patients' buy-in. Performance data included Milliman-reported data, information gained through regular calls and meetings with PMGs, and data reported to PBGH by PMGs.

Table II.A.6. PBGH self-reported IOCP encounter data

	Quarter							
	4 April-June 2013	5 July-September 2013	6 October-December 2013	7 January-March 2014	8 April-June 2014	9 July-September 2014	10 October-December 2014	11 January-March 2015
Total number of participants	443	1,329	2,823	4,337	5,586	6,571	7,267	7,770
Total number of encounters	507	1,678	3,434	12,179	25,987	19,898	24,485	23,758
In-person encounters	213	404	856	2,158	4,518	2,998	4,141	4,746
Telephone encounters	293	1,232	2,518	9,451	19,842	16,476	19,293	17,990
Other encounters	1	42	60	570	1,627	424	1,051	1,022
Percentage of total encounters								
In-person encounters	42	24	25	18	17	15	17	20
Telephone encounters	58	73	73	77	77	83	79	76
Other encounters	0	3	2	5	6	2	4	4

Source: Review of PBGH program reports, March 2015.

Note: The other encounters category includes encounters made by electronic communication or audio/video conference.

PBGH aimed to have 100 percent of participants complete a shared action plan at baseline, but did not meet this target. As of March 31, 2015, PMGs reported that 92 percent of participants had a shared action plan at baseline, although that percentage has increased since PBGH first reported the measure (62 percent) in the eighth quarter (April through June 2014).

PBGH recommended a target caseload of 125 participants per care manager. Many program staff reported preferring caseloads closer to 40 to 60 participants per care manager (or as low as 20 participants for less experienced care managers). Large caseloads especially burdened PMGs whose participants were distributed across a wide geographic region, making face-to-face visits time-consuming and sometimes impossible.

d. Staffing measures

By December 2014, PMGs had hired more than 214 full-time equivalent (FTE) program staff, exceeding PBGH's overall cumulative program staffing target of 211 FTEs. However, the ability to recruit and hire new program staff varied by site. One of the sites we visited could not find permanent, full-time care managers and instead relied on part-time temporary staff who often had little or no relevant care management experience. Another site we visited remained understaffed until 2015.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.7 summarizes the most significant facilitators and barriers in each of these domains influencing the effective implementation of the IOCP.

Table II.A.7. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Adaptability of program components • Perceived relative advantage of the IOCP program to deliver better quality care to high-risk participants with complex conditions • Feasibility of addressing the unique needs of each participant	• Feasibility of addressing the unique needs of each participant
Implementation process	• Engagement of physicians and practice staff	• Engagement of physicians and practice staff • Insufficient program staff to successfully recruit, enroll, and provide care management services to target number of participants
Internal factors	• Support from leadership to focus on providing good quality care to participants, not financial returns • Effective team communication with physicians and program staff	• Structural characteristics of IPAs create difficulties engaging physicians and accessing participants' information
External factors		• Wide range of EHR systems used by different providers • Lack of coordination with between providers who treat participants

Source: Review of PBGH program reports, March 2015.

a. Program characteristics

Two characteristics of the PBGH program helped PMGs implement the program: (1) adaptability of the program and (2) perceived relative advantage of the program compared with the standard delivery of care (Table II.A.6). First, program leaders attributed the program's success to the flexibility of its design. During the first year of participant enrollment, program administrators described how they tracked enrollment and met regularly with program staff to change and adapt the program to achieve program goals. For example, during the initial phase of implementation, some PMGs faced challenges reaching enrollment targets and implemented a variety of changes to increase enrollment, including expanding the target population, hiring part-time care managers, and recruiting participants in person instead of by telephone. During our second site visit, one program administrator stated that "being flexible and nimble was a huge

lesson for us," and advised other organizations implementing similar programs to "adapt and not get locked into one way of doing something."

Second, administrators, PCPs, and care managers perceived that the care managers improved the quality of care delivered to high-risk participants with complex conditions and reduced burden on PCPs, compared with the status quo way of delivering care to this population. PCPs interviewed described how care managers gave them helpful information that they otherwise would not have obtained, especially regarding psychosocial issues and physical barriers in participants' homes. Care managers were better positioned than PCPs to monitor participants' adherence to their care plans and check-in with them regularly, thereby offering dependable emotional support and practical health education. PCPs also commented that care managers were more likely to deliver timely and patient-centered services compared with other ancillary providers, such as home health agencies.

The ability of providers to address the needs of each participant both facilitated and challenged program implementation (Table II.A.7). Physicians and program staff agreed that care managers were well positioned to identify and resolve the complex social barriers participants face to adhere to their care plans, which supported program implementation. During our second site visit, one PCP stated, "[Care managers] have been able to reach patients or change behavior where I haven't been able to. They might have more time and more focus for that sort of thing, whereas I have to treat [medical conditions]." Similar to what we found during the April 2014 site visits, care managers continued meeting with participants on a face-to-face basis, typically at least once in each participant's home. They also facilitated the provider's understanding of each participant's unique circumstances and care needs. However, care managers found it difficult to meet the needs of certain participants. For example, care managers could not adequately communicate with participants with behavioral and mental health issues and severe cognitive impairments, and could not change the attitudes of participants who were unwilling to engage in their health care.

b. Implementation process

Across the three PBGH sites we visited, one implementation process factor both facilitated and challenged program implementation, and another factor only challenged implementation (Table II.A.7). First, respondents generally talked about the importance of physician engagement in making the program successful but also noted that physician engagement in the program varied. Interview respondents described multiple strategies to successfully engage physicians and practice staff who were not directly involved in implementing the IOCP. The most effective strategy involved embedding care managers in the primary care practices, meaning care managers were physically located in the practice for at least half a day each week. This facilitated engagement by providing opportunities for care managers to interact in person with physicians and practice staff, enabling everyone to familiarize themselves with one another and to build trusting relationships. As one physician commented about the embedded care managers during our second site visit, "Having their presence in the practice, seeing their faces, and knowing them personally makes a big difference. Then you know what kind of care your patients are getting." Lastly, to increase physician engagement, PMGs also proactively educated

physicians about the IOCP care management services, including presenting data that depicted improved outcomes for participants enrolled in IOCP.

Two of the sites described challenges resulting from underestimating the amount of program staff needed to successfully recruit, enroll, and provide care management services to participants. Recognizing this limitation, both sites increased program staff to overcome this implementation barrier. They also stopped conducting participant enrollment in participants' homes and solely enrolled participants during their hospital stays. One site hired more program staff and another recruited part-time, temporary care managers. However, the part-time, temporary care managers did not always have the appropriate skills to provide care management services and required extra oversight by management to match participants' needs to the skill levels of the care managers.

c. Internal factors

Characteristics of the organization implementing a program can influence implementation effectiveness. Two internal factors facilitated implementation of the PBGH program: support from leadership and communication between PCPs and care managers (Table II.A.7). First, PMG leadership expressed a commitment to sustaining the IOCP care management model despite the lack of short-term financial return. As one administrator stated during our second site visit, "We haven't done any analysis that looks at a return on investment. We do it because it is right for the patient." Second, frontline staff and physicians described how having care managers, clinical pharmacists, and social workers embedded in physicians' offices facilitated communication. Because physicians are not always linked to a central EHR system, face-to-face interactions with care coordinators is an efficient alternative. Physicians also expressed comfort communicating with care managers via email, text message, or telephone that resulted from personally knowing care managers. As one physician reflected during our second site visit, "If the communication wasn't sufficient, I can pick up the phone and call them. I know where they are, I can picture their faces and where they sit."

IPAs faced unique challenges to implementation, especially with regard to engaging physicians and identifying participants using health records. Engaging independent physicians proved challenging in some cases because they were often accustomed to working autonomously, had little or no in-person contact with the PMG, and were not obligated to respond to the PMG. "It's hard to develop the relationship and establish who you are," one PMG administrator observed during our second site visit. "We don't own the physicians' buildings or clinics, or hire the physicians' staff. Rather, the IPA is owned by the physicians." Finally, physicians in an IPA are not part of a centralized EHR, limiting the PMG's access to patients' data needed to identify participants and monitor enrollees.

d. External environment

Features of an organization's external environment can also influence implementation effectiveness. Two external factors presented challenges to the implementation of the program: participants' needs and resources and the technological environment. First, characteristics of the target population made enrollment difficult. During out second site visits, some program

administrators hypothesized that Medicare fee-for-service patients were more challenging to enroll than Medicare Advantage patients, because Medicare Advantage patients generally are more active in their health care. In an effort to expand enrollment to more patients who would benefit from care management services, some PMGs opened program eligibility criteria to include Medicare Advantage patients. Second, for care managers to succeed, providers needed access to comprehensive medical information on program participants. The wide range of EHR systems used by different providers, and the lack of coordination with other providers that treated participants, impeded the ability of care managers to access participants' information, particularly in the IPA setting. Lack of access to comprehensive participant information made it difficult for care coordinators to track changes in participants' conditions after follow-up physician visits and to continually assess participants' risk scores.

4. Sustainability and scalability

PBGH actively planned for the sustainability of the IOCP from the beginning of its award. In November 2014, PBGH shifted the focus of its technical assistance and training activities to sustaining the IOCP. It also organized the PBGH sustainability academy series (Table II.A.8). In addition, PBGH polled 20 of the 23 PMGs and found that many will sustain the IOCP model, in whole or in part, by maintaining all or most of the guardrails. All of the participating PMGs will continue using the shared action plan, 90 percent will continue the initial face-to-face super visits, 50 percent will integrate IOCP into other care management programs, 45 percent will continue using the participant activation measure, and some will transition to open source measures of participant engagement. In addition, some PMGs plan to expand the program to their commercially insured patient populations.

Table II.A.8. PBGH sustainability academy series

Sustainability academies	
Collaborative longitudinal academy	The academy catalyzed sustainability through didactic, experience-based collaboration, action planning, and longitudinal support toward actions.
Medical director leadership summit	Medical directors attended a one-day workshop that used didactic, peer-to-peer learning to promote strategic thinking for sustaining IOCP and similar primary-care based interventions for medically complex participants. Topics included changing their roles; demonstrating effectiveness; providing leadership through transition; managing organizational change; and defining, measuring, and achieving success in the evolving health care landscape.
Care managers sustainability academy	Care managers attended a session to discuss the next generation of care management, what it will look like, and how it will be operationalized. Topics included using data as a management tool, building high-performance teams, and doing more care management with fewer resources.

Source: Rev Review of PBGH program reports, March 2015.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Survey

Information gathered from interviews with program leadership and frontline staff at selected clinical sites or satellite offices provided important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff are limited to a small number of clinical locations and might not reflect the perspectives of clinicians practicing at other sites. To assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in fall 2014, the third year of the HCIA-funded program. The clinicians surveyed include PCPs (physicians, nurse practitioners, and physician assistants) but not the care managers themselves. Data from the survey provide additional insights into the implementation process and experience, as well as the contextual factors that might affect implementation effectiveness at PBGH.

In this section, we report on PBGH clinicians' views of their daily work life and practice. First, we focus on the contextual factors that can affect program implementation, including the characteristics of the practice locations, career satisfaction and burnout, barriers to providing high quality and patient-centered care, and clinicians' perceptions of how well the care teams function. We then present data on the alignment of PBGH's clinicians' views and experiences with the overall goals of the HCIA-funded program, as well as their awareness of and participation in the IOCP and their views of the barriers to and facilitators of successful program implementation.

2. Characteristics of clinicians and their practices

a. Characteristics of clinicians' practice locations

A total of 312 clinicians responded to the survey (resulting in a response rate of 64 percent). Of the respondents, 292 were physicians and 16 were mid-level providers (Table II.B.1).[1] These clinicians practiced predominantly at clinical locations with three or more clinicians (60 percent). PBGH clinicians reported that their primary source of compensation is a salary adjusted for performance (52 percent).

The clinicians reported working in settings that are advanced in terms of health IT (Table II.B.2). Although nationally, slightly more than half of physicians practice in settings with functional EHRs (Furukawa 2014), most clinicians reported using health IT at their practice locations. More than two-thirds of responding clinicians reported using electronic health systems for various functionalities, including use of electronic tracking systems and patient registries, advanced functions that are not in widespread use nationally (DesRoches 2014). Clinicians also reported that they offer patient-facing technologies such as prescription refill requests, appointment requests, and ability to email a clinician about a medical question or concern.

[1]The number of clinicians in each response category (here and throughout this section) does not always sum to the total number of PBGH respondents (N = 312) due to survey item nonresponse, as well as clinicians who reported that a given question did not apply to their practice and thus did not provide a response.

Table II.B.1. Types of clinicians, practices, and compensation sources

Survey item	Number of respondents	Percentage of Respondents
Type of clinician		
Physician	292	94%
Nurse practitioner or Physician assistant	16	5%
Type of practice		
Group practice (3 or more clinicians)	186	60%
Solo practice	41	13%
Group or staff model health maintenance organization	21	7%
Two-clinician practice	19	6%
Private or nonprofit hospital	15	5%
Federally Qualified or other community health center	15	5%
Other	11	4%
Primary compensation source		
Salary adjusted for performance	161	49%
Fee for service	54	17%
Fixed salary	53	17%
Other (hourly/time-based, other)	29	10%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

Table II.B.2. Electronic capabilities for clinicians and patients

Survey item	Number of respondents	Percentage of respondents
Physicians use EHR to		
Access laboratory results	289	93%
Enter clinical notes	288	92%
Receive drug dosing and interaction alerts	281	90%
Prescribe medications	278	89%
Order tests and procedures	268	86%
Access participant registries	224	72%
Track electronic referrals	222	71%
Patients can		
Email clinician about a medical question or concern	208	67%
Refill prescriptions	210	67%
Request appointments	190	61%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall careers, level of burnout, and perceptions of their practice environments can all affect the success of program implementation and organizational change. Clinicians are generally satisfied with their careers in medicine (Table II.B.3). However, only 31 percent reported being very satisfied and almost 28 percent of physicians were experiencing some symptoms of burnout at the time the survey was taken.

Table II.B.3. Career satisfaction and burnout

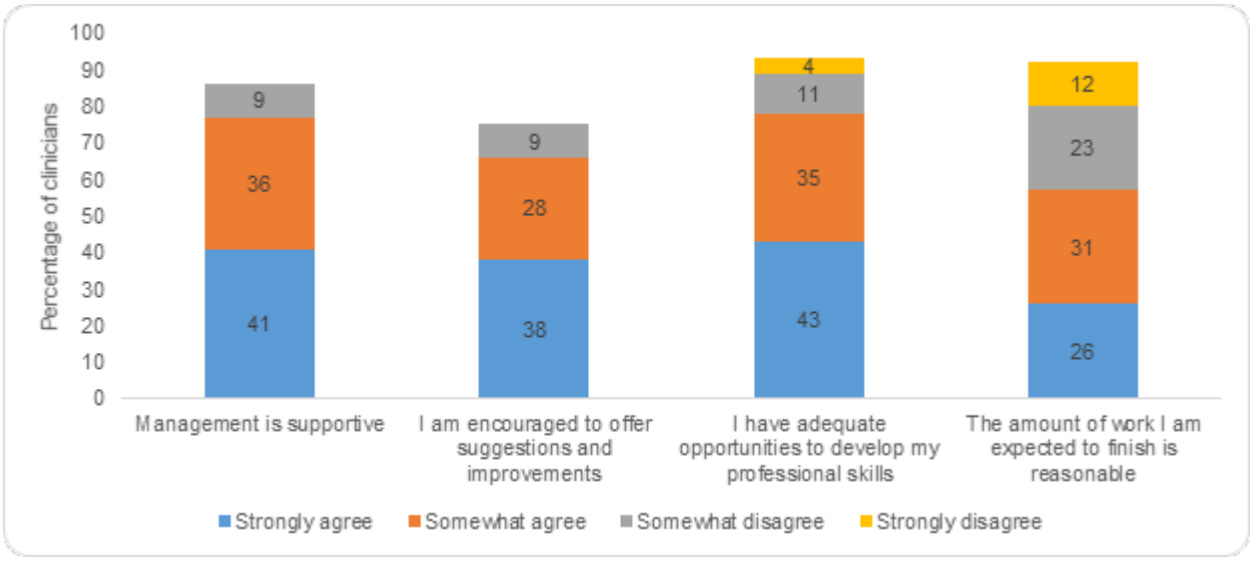
Survey item	Number of respondents	Percentage of respondents
Overall satisfaction with career		
Very satisfied	96	31%
Somewhat satisfied	141	45%
Neither	13	4%
Somewhat dissatisfied	45	14%
Very dissatisfied	11	4%
Degree of burnout		
I enjoy my work. I have no symptoms of burnout.	58	19%
Occasionally I am under stress, and I don't always have as much energy as I once did, but I don't feel burned out.	150	48%
I am definitely burning out and have one or more symptoms of burnout, such as physical and emotional exhaustion.	60	19%
The symptoms of burnout that I'm experiencing won't go away. I think about frustrations at work a lot.	27	9%
I feel completely burned out and often wonder if I can go on. I am at the point where I may need some changes or may need to seek some sort of help.	--	--

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

The clinicians gave similar ratings to their workplace management. Most responding clinicians either agreed or strongly agreed that their management team was supportive, that they were encouraged to offer suggestions and make improvements, and that they had adequate opportunities for professional development (Figure II.B.1). However, fewer than 60 percent of respondents agreed that the amount of work they were expected to complete each day was reasonable.

Figure II.B.1. Workplace ratings



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Totals may not add to 100% due to survey item non-response.

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. Slightly more than half of responding clinicians somewhat or strongly agreed with the statement "It is possible to provide high quality care to all of my patients" (Table II.B.4). For major barriers to providing optimal care, most clinicians reported lack of time to spend with patients, too many reminders on the EHR, insufficient reimbursement, patients' inability to pay for care, difficulties obtaining specialized diagnostic tests or treatments, and lack of timely information about care provided to patients by other physicians.

Table II.B.4. Perceptions of ability to provide high quality care

Survey item	Number of respondents	Percentage of respondents
It is possible to provide high quality care to all of my patients		
Strongly agree	68	22%
Somewhat agree	113	36%
Neither agree nor disagree	44	14%
Somewhat disagree	63	20%
Strongly disagree	16	5%
Percentage reporting each of the following at least somewhat limits their ability to provide optimal, patient-centered care		
I do not have enough time to spend with patients during visits.	253	81%
I receive too many reminders from my EHR.	184	59%
I lack timely information about the patients I see who have been care for by other physicians.	235	75%
I lack adequate information from research evidence to guide my clinical decisions.	98	31%
It is difficult for me to obtain specialized diagnostic tests or treatments for my patients in a timely manner.	161	52%
It is difficult for me to obtain specialist referrals for my patients in a timely manner.	130	42%
My patients have difficulty paying for needed care.	188	60%
The level of reimbursement is not adequate.	243	77%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Totals may not add to 100% due to survey item non-response.

c. Clinicians' perceptions of care team functioning

More than three-quarters (77 percent) of responding clinicians reported working as part of a care team and, overall, their perceptions of how these teams function were positive. Most clinicians working in care teams agreed that members of the care team relayed information in a timely manner (68 percent), had sufficient time for patients to ask questions (63 percent), used common terminology when communicating with one another (66 percent), verbally verified information they received from one another (57 percent), and followed a standardized method of sharing information when handing off patients (55 percent).

d. Alignment with goals of PCR

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The inclusion of the extremely important category helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are merely important. Most clinicians rated 8 of the 13 goals as extremely important (Table II.B.5).

Table II.B.5. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Reducing hospital readmissions	189	61%
Reducing ED visits	187	60%
Increasing access to primary care	180	58%
Improving care coordination for patients with chronic conditions	179	57%
Improving care continuity in primary care	173	56%
Improving patients' capacity to manage their own care	167	54%
Improving appropriateness of care	168	54%
Reducing overall health care spending	163	52%
Increasing the use of evidence-based practices in clinical care	136	44%
Improving the capability of health care organizations to provide patient-centered care	121	39%
Improving capability of health care organizations to provide team-based care	114	37%
Increasing the number of primary care practices functioning as a patient-centered medical home	105	34%
Increasing use of EHRs and other health IT	83	27%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

IT = Information technology, EHR = electronic health record

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the IOCP is to strengthen participating medical groups' capabilities of identifying medically complex participants and providing them with personalized care management services. Clinicians who believe personalized care management services could improve participants' outcomes might be more likely to refer medically complex patients from their panels. Alternatively, those who feel more negatively about the program might be less likely to refer patients. In this section, we report on clinicians' experiences with and perceptions of IOCP.

a. Awareness of the program and receipt of training

Only 31 percent of the clinicians we surveyed were at least somewhat familiar with the IOCP. Unfortunately, we cannot distinguish which survey respondents are part of IPAs versus integrated health systems. However, we know from both site visits most clinicians surveyed are part of IPAs in which clinicians practice independently from the PMG. IPAs often have unique barriers to communicating with physicians, because physicians of the same IPA have varying capacity for electronic communication and are dispersed in their physical locations. Another reason for the lack of familiarity might be because, in some cases, program leadership (from both

IPAs and integrated health systems) chose to integrate the IOCP program into existing clinical workflows without identifying it as a new program among clinicians. Clinicians did not receive training as part of IOCP implementation, although some might have participated in meetings to discuss the program and some have IOCP care managers embedded in their practices for at least half a day each week.

b. Perceived effect of program on participants' care

Clinicians' perceptions of the effect of the IOCP on the care they provide to participants were mixed. More than half of clinicians who were familiar with IOCP believed the program would have a positive effect on the quality and patient-centeredness of the care they provide, as well as on their ability to respond to participants' needs in a timely way (Table II.B.6). Fewer than half of physicians familiar with IOCP believed the program would have a positive effect on efficiency, safety, and equity. Very few clinicians perceived a negative impact of the program; rather, they believed the intervention would have no effect on the care they provide or that it was simply too soon to tell.

c. Barriers to and facilitators of program implementation

Finally, we asked the clinicians who were at least somewhat familiar with IOCP to rate the effect of a series of barriers to and facilitators of program implementation. The availability of community resources to care for participants with complex conditions and the availability of relevant information at the point of care were seen as having a positive effect on implementation by more than half of the surveyed clinicians (Table II.B.7). The other factors that clinicians perceived as having a positive effect were the availability of personnel and the quality of interpersonal communications with other allied health professions. Eighteen percent of respondents thought the amount of required documentation was a barrier (data not shown).

4. Conclusions about clinicians' attitudes and behaviors

The challenge of engaging PCPs, specifically in the IPA model, may be reflected in the low percentage of clinician respondents to the HCIA-Primary Care Redesign Clinician Survey. However, those who responded to the survey generally agreed that the program had a positive effect on the quality and patient-centeredness of the care they provide. The views of responding clinicians also tended to align with the goals of primary care redesign. Unfortunately, the results of this survey are not necessarily representative of all clinicians participating in the IOCP.

Table II.B.6. Perceptions of effects of program on participants' care

Survey item	Positive effect		Negative effect		No effect		Too soon to tell	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Perceived effect of the HCIA program on the care they provided to participants over the past year, including on:								
Quality of care	52	52%	--	--	13	13%	32	32%
Ability to respond in a timely way to participants' needs	53	53%	--	--	18	18%	25	25%
Efficiency	41	41%	--	--	21	21%	32	32%
Safety	44	44%	--	--	18	18%	33	33%
Patient-centeredness	53	53%	--	--	14	145	28	28%
Equity of care for all participants	32	32%	--	--	26	26%	37	37%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Figures are based on the total number of PBGH clinicians reporting they were at least somewhat familiar with the IOCP program. Data are not presented when the number of responses is fewer than 11 because of confidentiality restrictions.

Table II.B.7. Barriers to and facilitators of program implementation

	Positive impact		No impact		Not applicable/don't know	
Survey item	Number	Percentage	Number	Percentage	Number	Percentage
Level of program funding	33	33%	13	13%	44	44%
Amount of required documentation	22	22%	24	24%	31	31%
Availability of personnel	46	46%	15	15%	27	27%
The amount of time required by the program	28	28%	26	26%	34	34%
Availability of relevant participant information at the point of care	50	50%	14	14%	28	28%
Required use of computer and communications technology	38	38%	20	20%	30	30%
Availability of evidence-based clinical information	37	37%	21	21%	36	36%
Availability of community resources to care for participants with complex conditions	53	53%	16	16%	26	26%
Quality of interpersonal communications with other providers	43	43%	17	17%	29	29%
Quality of interpersonal communications with specialists	39	39%	22	22%	32	32%
Quality of interpersonal communications with other allied health professionals	47	47%	20	20%	26	26%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Figures are based on the number of clinicians who reported being at least somewhat familiar with the IOCP. Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Eighteen percent of respondents believe the amount of required documentation negatively impacted program implementation. Data are not shown for other survey items that respondents believe negatively impacted program implementation because there were fewer than 11 respondents.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

PBGH received HCIA funding to work with 23 PMGs across five states to implement the IOCP. The IOCP provided care management and care coordination services to Medicare beneficiaries with chronic conditions and/or risk of experiencing a hospitalization. The program aimed to reduce total health care spending and utilization while improving quality. Although PBGH faced delays with generating risk-stratification lists to support PMGs in identifying participants, the IOCP's flexible program design enabled PMGs to develop and implement alternative strategies to identify participants. As a result, PBGH avoided any delays to its program implementation time line and facilitated PMGs' ability to reach their enrollment targets. Compared with integrated health systems, IPAs generally faced more challenges during the implementation process. The independence with which PCPs function in the IPA model made it difficult for IPA leadership to engage PCPs, forcing many IPAs to rethink their physician engagement strategies after the program had been launched. Results from the HCIA-Primary Care Redesign Clinician Survey showed that only 31 percent of respondents were familiar with the IOCP, which is partially attributable to the challenges IPA leadership faced when

communicating with physicians. Of the clinicians who were familiar with the IOCP, most believed that the program had a positive effect on the quality of care, patient-centeredness, and their ability to respond in a timely way to participants' needs.

Our next steps for this evaluation are to (1) monitor PBGH's ongoing program implementation and any plans for sustaining the program beyond the funding period by reviewing quarterly data submitted by PBGH and (2) evaluate trainees' and clinicians' attitudes and experiences with the program in the third year of the award through administered surveys. We are currently unable to assess whether and how the program affected participants' outcomes due to limitations in identifying a comparison group.

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

REFERENCES

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low. *Health Affairs,* vol. 33, no. 9, 2014, pp. 1672–1679.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

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

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for PeaceHealth Ketchikan Medical Center

March 2016

Boyd Gilman
Purvi Sevak
Victoria Peebles
Greg Peterson
Catherine DesRoches
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

PEACEHEALTH KETCHIKAN MEDICAL CENTER

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by PeaceHealth Ketchikan Medical Center (PeaceHealth) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the PeaceHealth program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on patients' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF PEACEHEALTH

PeaceHealth received a three-year, $3.2 million award to implement the Better Health Through Coordinated Care—A Plan for Southeast Alaska program (hereafter referred to as the coordinated care program) within two primary care practices located in island communities in southeastern Alaska. Table I.1 summarizes key features of the program. The program involved four interrelated components: (1) transitional care services for all patients discharged from the PeaceHealth Ketchikan Medical Center and intensive transitional care services for patients with congestive heart failure (CHF) who are on a PeaceHealth provider panel (paneled patients are defined as patients seeking usual care with a PeaceHealth provider); (2) short-term care management for patients with a temporary medical or social hurdle; (3) long-term case management for patients requiring assistance to effectively manage their chronic conditions; and (4) population health management, including redefining the scrub-and-huddle process and outreach to paneled patients to improve preventive care.

The goals of the program were to (1) improve access to primary care by hiring staff and increasing after-hours care; (2) increase support to and improve outcomes for high-risk patients (particularly those with diabetes, heart failure and, later, hypertension and high-risk pregnancies) by hiring care coordinators to manage chronic conditions and link patients to community resources; and (3) strengthen primary care teams by enhancing the skills of medical assistants and implementing routine scrub-and-huddle procedures. Through its coordinated care program, PeaceHealth aimed to reduce unplanned hospital admissions, emergency department (ED) visits, and average total cost of care, particularly among patients with CHF, diabetes, and hypertension.

Table I.1. Summary of PeaceHealth Ketchikan Medical Center PCR program

Awardee's name	PeaceHealth Ketchikan Medical Center
Award amount	$3,169,386
Implementation date	October 18, 2012
Award end date	June 2015
Program description	1. **Transitional care.** Care coordinators contact all patients discharged from PeaceHealth Ketchikan's ED or hospital who are on a PeaceHealth provider panel. 2. **Short term-care management.** Primary care providers refer patients to the care coordinators for short-term social and behavioral health needs. 3. **Long-term case management.** Care coordinators work with patients who have chronic diseases for three or more encounters to manage their conditions. 4. **Population health management.** Care coordinators identify and reach out to patients with uncontrolled chronic conditions (diabetes, CHF, hypertension and, later, high-risk pregnancies).
Innovation components	Care transitions, care coordination, care management, population health
Intervention focus	Practice-level
Workforce development	1. Created new care coordinator positions and trained staff through a care coordination program offered through Oregon Health and Sciences University 2. Created an internal training program for MOAs and developed a new online certification program for MOAs through the University of Alaska
Target population	1. All patients discharged from PeaceHealth Ketchikan Medical Center who are paneled to a PeaceHealth provider 2. Patients with chronic conditions (including diabetes, CHF, hypertension, and high-risk pregnancies) who are paneled to a PeaceHealth provider
Program setting	Provider-based (primary care practices)
Market area	Local (Ketchikan and Craig, Alaska)
Market location	Rural (remote island communities in southeastern Alaska)
Core outcomes	• 20 percent reduction in 30-day hospital readmission rates • 75 percent reduction in ED costs • 15 percent reduction in total costs

Sources: Review of PeaceHealth program reports and information collected on site during site visits in May 2014 and April 2015.

CHF = congestive heart failure; ED = emergency department; MOA = medical office assistant.

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the program and its components, highlighting how it has been adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external factors. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of PeaceHealth's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collecting during site visits conducted in May 2014 and April 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

The PeaceHealth coordinated care program included four components: (1) transitional care, (2) short-term care management, (3) long-term case management, and (4) population health management (Table II.A.1).

b. Target populations and patient identification, recruitment, and enrollment

As Table II.A.1 shows, the target population varied by component: short-term care management and long-term case management primarily focused on patients with diabetes, CHF, hypertension, and high-risk pregnancies, whereas the transitional care and population health components were available to all patients.

Patient identification strategies for the PeaceHealth program varied by component. For the transitional care component, program staff used hospital discharge data to identify patients discharged in the previous 24 hours from the PeaceHealth Ketchikan Medical Center hospital. Although program staff called all patients on the list, they stratified discharges into three groups (red, yellow, and green), indicating their risk of rehospitalization, based on demographic and diagnostic information available from their medical records. Stratification characteristics included the following:

- **Demographics.** Age and race
- **Prior hospital admission.** Two admissions in the past year, one admission in the past 180 days with a length of stay of three or more days, or one admission in the past 30 days
- **Diagnosis.** End stage renal disease, chronic obstructive pulmonary disease, diabetes, and mental health
- **Medications.** At least five active prescription medications
- **Charlson Comorbidity Index.** Ten-year mortality rate based on several comorbid conditions
- **Receipt of charity care.** Eligibility for Bridge Assistance, a PeaceHealth financial assistance program
- **ED visit.** Co-occurring visit to the hospital ED

PeaceHealth developed the risk score on the hospital discharge form based on a risk-stratification analysis completed by Whatcom Alliance for Health Advancement, using data from August 1, 2012, to July 31, 2013.

Table II.A.1. Key details about program design and adaptation

	Program component			
	Transitional care	Short-term care management	Long-term case management	Population health management
Target population	Program staff targeted all patients discharged from the PeaceHealth Ketchikan Medical Center and on a PeaceHealth provider panel, and offered enhanced transitional care services for patients with CHF.	Short-term care management targeted patients with diabetes, CHF, and later hypertension and high-risk pregnancies. Short-term social work case management services were also available to any patient with an identified psychosocial need.	Program staff targeted patients with diabetes, CHF, and later hypertension and high-risk pregnancies.	Outreach calls targeted diabetic patients. All patients benefitted from the scrub-and huddle process. Scrubbing involved reviewing a patient's medical records to identify outstanding care needs, such as laboratory tests, mammograms, immunizations, or colorectal screenings, and the huddling process was a team meeting to review a patient's needs before a regularly scheduled visit.
Patient Identification	Program staff identified patients listed on the PeaceHealth Ketchikan Medical Center hospital discharge form in the previous 24 hours.	Program staff identified patients before appointments based on a diagnosis of diabetes or CHF, and later expanding to hypertension and high-risk pregnancies. Physicians also used their clinical judgement when deciding whether to refer someone to case management for psychosocial issues.	Program staff identified patients identified based on a diagnosis of diabetes or CHF, and later expanded to hypertension and high-risk pregnancies.	Program staff identified patients via a diabetic outreach report (patients without scheduled diabetic follow-up appointments).
Patient recruitment and enrollment	Patients were not actively recruited and enrolled into the program. Program staff considered patients in the target population for each component enrolled in the program.	Patients were not actively recruited and enrolled into the program. Program staff considered patients in the target population for each component enrolled in the program.	Patients were not actively recruited and enrolled into the program. Program staff considered patients in the target population for each component enrolled in the program.	Patients were not actively recruited and enrolled into the program. Program staff considered patients in the target population for each component enrolled in the program.

Table II.A.1 *(continued)*

	Program component			
	Transitional care	**Short-term care management**	**Long-term case management**	**Population health management**
Service delivery protocol	A daily patient discharge protocol guided program staff through telephone follow-up after hospital or ED discharge. Staff reviewed discharge instructions, assessed the need for additional support or patient education, and ensured patients understood their medications. Patients with CHF also received telephone calls 14 and 28 days after discharge to ensure they made a follow-up appointment with their PCP and reviewed medications and signs of fluid volume excess.	No formal documented protocols were in place.	No formal documented protocols were in place.	Protocols for diabetic outreach and health maintenance were used. Diabetic outreach reports identified patients for outreach and standardized documentation of these calls. MOAs completed health maintenance worksheets to help providers prepare for patients' visits.
Adaptations	Initially targeted only high-risk patients, and later expanded to include all discharges (with less intensive telephone follow-up for patients not at high risk).	The target population was expanded to include patients with hypertension and high-risk pregnancies.	The target population was expanded to include patients with hypertension and high-risk pregnancies.	This component initially focused on overdue mammograms and colorectal cancer screenings, uncontrolled high blood pressure, and a positive tobacco status with no counseling. It later expanded to other conditions and screenings.

Sources: Interviews from second site visit, April 2015; document review, March 2015.

CHF = congestive heart failure; MOA = medical office assistant; PCP = primary care physician.

To identify patients for the short-term care management and long-term case management components of the HCIA program, staff used an implicit risk score assessment, focusing initially on diabetes and CHF, and later expanding to hypertension and high-risk pregnancies. Physicians also used their clinical judgement when deciding whether to refer a patient to case management for psychosocial issues. Patients were identified for the population health management component through the diabetic outreach report, which identified patients without diabetic follow-up appointments. Program staff performed the scrub-and-huddle process for all patients. Scrubbing involved reviewing a patient's medical records to identify outstanding care needs, and the huddling process was a team meeting to review a patient's needs before a regularly scheduled visit. Some care teams did a less formal huddle because they frequently discussed patients throughout the day; others went through a formal scrub and huddle at the beginning of each week or day before seeing their patients.

PeaceHealth did not formally recruit or enroll patients into the program. Rather, program staff considered all patients in the target population for each component eligible for and enrolled in the program. Patients often knew they were referred for additional services or contacted about making follow-up visits, but they typically would not have been aware that they were enrolled in a particular program to receive services.

c. Service delivery protocols

For the transitional care component, care coordinators used a defined protocol for making post-discharge follow-up telephone calls, with a more detailed protocol used when making follow-up telephone calls for patients with CHF. The protocol prompted care coordinators to assess the patient's status, review medications and current symptoms, order any needed equipment, and schedule follow-up appointments. Patients with CHF received additional follow-up calls 14 and 28 days after discharge to review their weight, assess symptoms of fluid volume excess, and review medications. The protocol for daily patient discharge calls prompted care coordinators to ask how patients were doing, review symptoms to watch for and identify red flags, review medications, help patients schedule follow-up appointments, and help patients obtain the equipment they needed. Patients with CHF received additional follow-up telephone calls 14 and 28 days after being discharged from the hospital or ED.

For the short-term care management and long-term case management components, PeaceHealth did not use formal service delivery protocols. Care coordinators typically responded to the instructions provided by the clinicians and developed plans of action based on an individual patient's needs and the specific skill set of the care coordinator. When asked if protocols would have been helpful, care coordinators responded that patients had too many unique needs, so creating general protocols would not have been desirable or feasible.

The population health management component of the program included protocols for diabetic outreach and health maintenance. For diabetic outreach, medical care coordinators were instructed to run a report in EpicCare, an electronic health record system (EHR), the first week of every month. This report focused on patients with A1c levels greater than 8 and no follow-up appointment scheduled, low-density lipoprotein levels greater than 100, elevated blood pressure greater than 140/90, and no appointment scheduled in the next 30 days. Medical care

coordinators called patients meeting these criteria and documented the calls. They also documented whether a follow-up appointment had already been scheduled or, if appropriate, why a follow-up appointments had been missed. Medical care coordinators pulled this extract at the beginning of every month and worked through the list of patients as they had time. The program manager hoped to add a new tool that would require less data manipulation from the nurses, and would show in real-time when a patient was last seen, easing the administrative burden on care coordinators.

The population health management component included a health maintenance worksheet that medical office assistants (MOAs) completed as part of their role in the scrub-and-huddle process. Before a patient arrived for a visit, MOAs entered information in the health maintenance worksheet about diabetes, hyperlipidemia, hypertension, hypothyroid, well adult, colonoscopy, pap smear, mammogram, bone density (DEXA) scan, influenza, pneumococcal 23, human papillomavirus, tetanus-diphtheria, and tetanus-diphtheria-pertussis. In addition, PeaceHealth added space at the top of the worksheet to capture the patient's chief complaint, blood pressure, temperature, oxygen count, pulse, and physical examination. MOAs used this information to identify whether patients needed lab work or other health maintenance examinations before their next clinic appointment. The MOA communicated this information to the provider during the daily or weekly team huddle.

d. Intervention staff and workforce development

Program services were administered primarily through two positions: care coordinators and MOAs (Table II.A.2). The award created positions for four medical care coordinators, one social work care coordinator, and training for existing MOAs.

Care coordinators supported all four program components. Most providers felt that, to be effective in their position, care coordinators had to be registered nurses—three of the four medical care coordinators were registered nurses (the fourth was a licensed practical nurse). All medical care coordinators completed training on motivational interviewing, chronic illness, nursing assessment planning, and communication through courses offered by the Oregon Health and Sciences University. The social work care coordinator had recently become licensed and could provide long-term counseling, but her role on the HCIA-funded program focused on short-term care management; she did not provide regular or ongoing counseling services as part of the long-term case management program.

Several staff mentioned that, to fully address patients' needs, care coordinators required a hybrid of clinical expertise and social work experience. For example, medical care coordinators not only focused on clinically oriented nursing tasks, but were also responsible for identifying the underlying psychosocial issues preventing a patient from accessing and remaining in care or managing his or her conditions. Medical care coordinators also had to understand diabetic lab results and have the skills to teach patients to use diabetic supplies, to help them understand why managing their sugar was important, and to follow-up with appropriate support and education. The social work care coordinator also had to have some clinical knowledge to understand the medical challenges patients faced. PeaceHealth did not create the MOA position as part of the coordinated care program. Rather, the award provided funding to increase the competencies

Table II.A.2. Key details about staff involved in the coordinated care program

	Staff responsibilities by program component			
Staff position	**Transitional care**	**Short-term care management**	**Long-term case management**	**Population health**
Medical care coordinators	Follow-up with patients 1 or 2 days after discharge	Work with patients to manage their chronic conditions, including providing education or offering diabetic supplies	Work with patients to manage their own chronic conditions, including providing education or offering diabetic supplies	Contact patients with chronic diseases and helping them schedule appointments in the clinic
Social work care coordinator	Provide transitional care services if the discharged patient had behavioral health issues	Help patients apply for Medicaid, understand their insurance, coordinate transportation, connect them with other mental health resources, or assist with other psychosocial issues	None	None
Medical office assistant (MOA)	None	None	None	Conduct the scrub-and-huddle process Receive additional training to enable all staff to work toward the top of their licensure

Sources: Interviews from second site visit, April 2015; document review, March 2015.

of existing MOAs and help the care team work at the top of their licensure. MOAs previously were responsible for rooming patients and preparing them for the provider's exam. As part of the program's population health management component, MOAs were trained to execute the scrub-and-huddle process with the provider team. Training for the MOAs focused on six competencies: (1) point-of-care testing, (2) patient visit facilitation, (3) infection control, (4) medication administration, (5) exam room preparation, and (6) patient safety. The clinical educator also conducted monthly brown bags on topics such as immunizations or heart health and sent daily interesting facts to MOAs on educational topics, such as the definition of cholesterol or abnormal types of respiratory patterns.

As part of the award, PeaceHealth attempted to create a new medical assistant certificate training curriculum and certification program through the University of Alaska. The program is scheduled to start in September 2015 (three months after the end of the award). It plans to offer online courses for four semesters, include two weeks of hands-on training, and require 280 clinical hours to graduate. MOAs are not currently recognized in the state of Alaska, but this training would increase the skill sets of MOAs, enabling them to enter orders and helping PeaceHealth move from CMS Stage 2 to Stage 3 of meaningful use. At the time of our site visit, 7 of the 19 MOAs had expressed interest in enrolling in the new certification program.

In addition, PeaceHealth hired a nurse practitioner, clinical educator, and program coordinator to support the program's objectives. The nurse practitioner was hired to expand clinic hours. The clinical educator facilitated training among the MOA, and the program coordinator was hired to administer the program. None of these positions involved the provision of the four components of the coordinated care program.

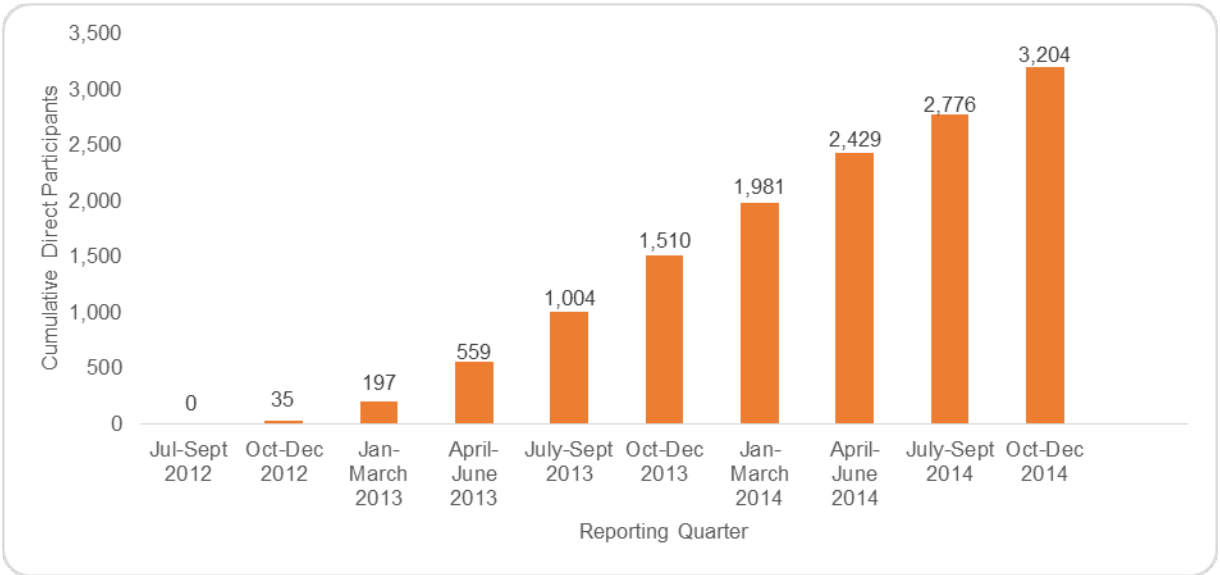
2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness. For this analysis, we rely on information collected during interviews with program administrators during the second site visit, as well as self-reported information included in PeaceHealth's quarterly self-monitoring and measurement reports.

a. Program enrollment

PeaceHealth's target population included all patients on a PeaceHealth medical group primary care provider panel in southeastern Alaska, but as stated earlier, PeaceHealth did not actively recruit and enroll patients into the program. Therefore, enrollment counts provided are based on patients in the target population who actually received services from an HCIA-funded position, such as a care coordinator or social worker. Across all four components, the program aimed to provide HCIA-supported services to a cumulative total of unique 3,500 patients by the end of the program's third year. By December 2014 (six months before the end of the program), PeaceHealth had achieved 92 percent of this goal, providing HCIA-funded services to 3,204 unique paneled patients since inception (Figure II.A.1).

Figure II.A.1. Cumulative number of unique direct participants, by program quarter



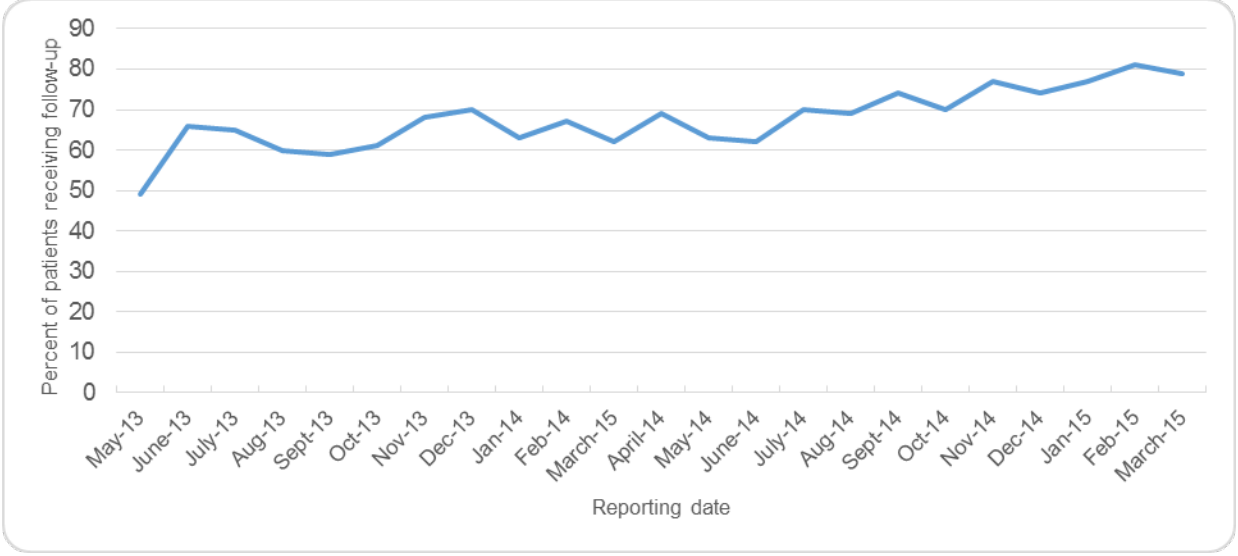
Source: Review of PeaceHealth's program reports, June 2015.

b. Service-related measures

PeaceHealth's quarterly self-monitoring and measurement reports (used by awardees to monitor performance and make improvements when needed) contained no information on patient encounters, so it was difficult to assess the implementation effectiveness of service delivery. Moreover, the systemwide adoption of EpicCare prevented the awardee from reporting such implementation metrics as the percentage of patients with hemoglobin A1c greater than 8 who had a follow-up call by a care coordinator.

PeaceHealth did report that by March 2015 nearly 80 percent of all patients discharged from the Ketchikan hospital received a follow-up call from a care coordinators for transitional care assistance (Figure II.A.2). It also reported a total of 1,672 patient encounters with a care coordinator during the 10th quarter of the program (October through December 2014), the latest period for which such information is available. Each encounter was counted separately so that an individual program participant could have received more than one encounter type in a single day or within a month or quarter. Of these 1,674 encounters, 1,384 were by telephone, 278 were in-person, and 10 were via telehealth or telemedicine technology.

Figure II.A.2. Percentage of hospital discharges receiving a follow-up call from a care coordinator



Source: Review of PeaceHealth's program reports, June 2015.

c. Staffing measures

PeaceHealth hired 9.5 full-time equivalent (FTE) staff members with HCIA funding, exceeding its staffing goals of 7.5 FTEs. Of these new recruits, 4.5 FTEs were for medical or social work care coordinators (see Table II.A.2 for a description of the care coordinator positions). PeaceHealth hired two additional care coordinators in the last year of the program with HCIA funding, one serving women with high-risk pregnancies and the other helping to address financial barriers to care. As reported, PeaceHealth also hired a nurse practitioner,

clinical educator, and program coordinator, each allocated 1.0 FTE, to support the program. PeaceHealth filled all of these positions as planned.

PeaceHealth trained the care coordinators and MOAs. However, it is still working to operationalize the new MOA certification program through the University of Alaska–Southeast. By the 10th program quarter (December 2014), care coordinators had completed 56 hours of care management training and the MOAs had completed 90 hours of training. PeaceHealth projected it would train a total of 84 staff members. The awardee exceed this goal, training 97 staff members by the end of the program in June 2015.

d. Program time line

Program administrators reported initial delays in implementing the program, affecting all of the components. It took several quarters for the program to become operational, only after it had hired staff, transitioned to the new EHR system, created reports, and developed protocols. The program became partially operational by January 2013 (six months after the award). However, it took another year to fully implement the transitional care component of the project. The implementation of the short-term care management and long-term case management components was delayed when one of the medical care coordinators resigned. By September 2013 (about one year after award), all of the program staff were in place. Finally, the population health management component experienced significant delays in creating the medical assistant training program through the University of Alaska due to staff turnover and delays coordinating with the university.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external factors. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.3 summarizes the major facilitators of and barriers to PeaceHealth's implementation effectiveness in each domain.

a. Program characteristics

Three characteristics of the PeaceHealth initiative facilitated implementation of the program: (1) perceptions of the relative advantage of the program compared with the standard delivery of care, (2) frontline users' flexibility in implementing the program, and (3) adaptation of the program to meet patients' and providers' needs.

Although obtaining staff buy-in was initially a challenge, as the program continued to develop administrators, providers, and staff began to agree that the new model offered an advantage for improving care compared with the standard delivery of care. Frontline staff reported that, before the implementation of the PeaceHealth program, patients would come in for follow-up diabetes appointments, and providers would not have the necessary lab test results to assess their health, requiring patients to return for an additional visit after providers received and reviewed their lab results. With the new scrub-and-huddle process, MOAs identified and scheduled lab tests before the appointment, making the visit more effective for both the provider

Table II.A.3. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Perceptions of the relative advantage of the program compared with the standard delivery of care • Frontline users' flexibility in implementing the program • Adaptation of the program to meet patients' and providers' needs	• No major barriers identified
Implementation process	• Dedicating resources to support the program • Monitoring progress to guide ongoing improvement	• Engagement of and buy-in from staff
Internal factors	• Culture of the organization • Team collaboration • Structural characteristics of the two clinics	• No major barriers identified
External environment	• No major facilitators identified	• Technological environment • Environmental challenges of being in a remote location in Alaska

Sources: Interviews from second site visit, April 2015; document review, March 2015.

and patient. In addition, providers reported that they previously had to conduct six or seven office visits with each patient annually to address chronic conditions. As part of the care coordination program, care coordinators followed up with patients and shared information about their progress with providers between visits, reducing the workload for providers. Administrators also believed care coordinators offered an added value to the patients. For example, program staff said that patients were often overwhelmed with information in the hospital, and having someone connect with them at home after an inpatient or ED visit helped ensure they understood their medications and discharge instructions. Adding the social worker in the second year of the program was also seen as critical, due to the high number of psychosocial issues affecting patients' care that would otherwise go unaddressed.

Second, PeaceHealth's program gave frontline staff flexibility in implementing the care coordination model to meet an individual patient's needs and to address providers' preferences. For example, program administrators recognized the need for flexibility in administering protocols for care management and case management and showed a willingness to allow the team to think innovatively. Care coordinators were able to schedule their own appointments, customize services for patients, provide patients with access to educational materials and resources, and follow-up with them by telephone. There was also flexibility in how providers executed the scrub-and-huddle process: some care teams did a less formal huddle (because they shared the same office and frequently discussed their patients), whereas others went through a formal scrub-and-huddle at the beginning of each week or day before seeing their patients.

Third, PeaceHealth adapted its program to focus on specific populations and realign staff roles as the program learned more about workflows that were effective for particular groups of patients. The program initially focused its transitional care component on all discharges from the PeaceHealth Medical Center. It later narrowed its focus to include only those patients with CHF and diabetes because program leadership believed those patients could benefit the most from transitional care services. Later, the program shifted again to provide transitional care to all patients on a PeaceHealth panel who were discharged from the local hospital. The short-term care management component also originally focused on smoking cessation, but shifted its focus to patients with diabetes, and then added CHF, hypertension, and high-risk pregnancies. Program administrators, working with providers, determined that high-risk pregnancies were expensive for the PeaceHealth system, and there was a need in the community for these services.

b. Implementation process

Two implementation process factors facilitated the implementation of the care coordination program: (1) dedicating resources to support the program and (2) monitoring progress to guide ongoing improvement.

First, program leaders invested and focused HCIA resources toward areas they believed could have the biggest impact, using HCIA funding to hire new staff to provide care coordination and social work services, and to train MOAs and care coordinators. Recognizing that providers did not have time to address all of their patients' needs (such as diabetes education, routine health maintenance follow-up, and transitional care after hospitalization), PeaceHealth created the medical care coordinator position. In addition, PeaceHealth hired a social work care coordinator to addresses patients' psychosocial issues and link them to community resources. In the final year of the program, leadership identified two new target areas for investing resources: (1) helping patients to understand billing questions and address financial concerns and (2) providing care coordination services to patients with high risk-pregnancies. The leadership team hired two new care coordinators to address these issues. PeaceHealth developed training sessions to increase the competency level of existing MOAs, enabling nurses to work closer to the top of their licensure.

Second, PeaceHealth established a self-monitoring process that guided ongoing program revisions and quality improvement efforts. The self-monitoring process supported implementation by enabling PeaceHealth to make real-time adjustments to the program based on evidence. Program administrators tracked monthly patient enrollment, health maintenance, chronic disease, and quality-of-care process and outcome measures, and reviewed these data on a weekly basis to adjust operational plans as necessary. For example, program administrators noticed that hemoglobin A1c measures for patients with diabetes were rising. Because the chart review can be more complex for these patients, program administrators transferred the responsibility for the scrub of these patients from MOAs to care coordinators. Later, after the MOAs received additional training, the responsibility of diabetic scrubs was transferred back to MOAs.

Initially, staff engagement and buy-in was a barrier to implementation. For example, providers at first did not understand the purpose or role of care coordinators and were not sure

how to use them. To address this challenge, the program manager and clinical educator attended providers' meetings to clarify the role of the care coordinator. In addition, medical care coordinators asked for physicians' input when creating new protocols, moved their desks closer to providers' offices, and provided timely feedback to clinicians. The social work care coordinator created and circulated a pamphlet explaining her role and how she could assist the primary care team. She also reported following through with referrals and giving feedback to providers to close the loop on patients. Staff consistently said that the social work care coordinator was critical to the care team. After the care coordinator for high-risk pregnancies was hired, providers became involved in setting the goals of the new position, which facilitated earlier buy-in.

Some providers were also hesitant to buy into the scrub-and-huddle process. Although some doctors viewed this process as enabling them to focus on clinical care and make their appointments more thorough, some found that the appointment was inefficient because the necessary chart preparation was not conducted beforehand or they had to rework the chart preparation themselves if the MOA was not adequately trained and the scrub-and-huddle was not conducted properly. The frequency and timing of the huddle often depended on the provider and MOA relationship and varied across teams. Staff reported that the health maintenance worksheet and the additional training for the MOAs helped make the huddles more productive. MOAs felt they better understood the meaning of the lab results, the target values for these results, and the timing of ordering them. They also better understood when it was appropriate to ask the doctor to place the order.

c. Internal factors

Three internal factors facilitated implementation of the PeaceHealth program: (1) the culture of the organization, (2) team collaboration, and (3) the structural characteristics of the two practices.

First, PeaceHealth's corporate culture was a factor in deciding to apply and helped facilitate the program's implementation. Program staff said the intervention was consistent with PeaceHealth's mission and its overall approach to care. The alignment of goals between the program and the corporate office facilitated program leadership, despite the potential loss of hospital revenue from lower inpatient and ED service use.

Second, after overcoming initial concerns about the new care coordinator positions, frontline staff reported that care teams worked well together. Providers expressed confidence in care coordinators and trusted that care coordinators would communicate with them about patients' care when necessary. Providers reported that they were pleased with the support provided by the care coordinators.

Finally, the program operated within a larger health care system, which facilitated PeaceHealth's ability to implement the intervention. Because the participating practices and the hospital were part of the same corporate structure, discharge notifications and collaboration between the hospital and clinics were fairly streamlined. The care coordinators could easily access the daily hospital reports with a list of patients discharged from the hospital and the ED.

In addition, because there were few other providers on either island, the care coordinators were usually able to capture most of the relevant medical information for their patients.

d. External environment

Two external factors presented challenges to implementing the initiative at PeaceHealth: (1) the technological environment and (2) the remote location in Alaska.

First, independent of this HCIA initiative, PeaceHealth launched a new EHR system, called EpicCare, in September 2013 in the two clinics. However, the hospital still used its legacy EHR system (called Centricity) and did not plan to transfer to EpicCare until May 2016. Because the hospital and clinics used separate EHR systems, certain shared reporting and tracking functionality features were lost. However, the switch to EpicCare also facilitated several process improvements within the clinics. For example, EpicCare had the ability to extract information from patients' charts, which made the scrubbing and chart review processes easier. In preparation for the implementation of EpicCare, clinics also reexamined and standardized some of their workflow processes. This led to greater standardization and alignment of roles and responsibilities of MOAs with other practice staff. The clinics also considered purchasing wraparound software (called Healthy Planet) to support medical care coordinators by creating live-feed dashboards with actionable information from their administrative list of discharges.

Second, PeaceHealth's location on two islands in southeastern Alaska presented unique challenges for implementing this award. Both participating practices are located in remote areas, where it is often difficult to recruit and hire staff. PeaceHealth administrators struggled to fill the program manager position and experienced turnover among care coordinators in the first year. In addition, because the medical assistant role was not recognized as a credentialed medical position in Alaska, staff hired as MOAs often lacked formal training and had limited clinical skills (such as giving injections), compared with licensed medical assistants in other states.

4. Sustainability

PeaceHealth hoped to be able to continue funding the care coordinator positions after the end of the award, but at the time of our site visit, the awardee did not yet have a firm plan in place for covering the cost of these positions. The program leadership team met with officials from the state Medicaid program, but they were not optimistic about receiving additional reimbursement to cover the cost of the services due to the major financial constraints on the state budget due to a decline in oil revenues.

The program manager identified four possible funding streams to support these positions. First, PeaceHealth's corporate office has a grants procurement officer who monitors new grants and looks for additional funding opportunities and, at the time of our visit, PeaceHealth was exploring new grants (such as the CMS Transforming Clinical Practices Initiative) to fund portions of the program. Second, the corporate office was exploring the feasibility of using the Transitional Care Management codes available under Medicare fee-for-service billing to cover the costs of medical care coordinators. PeaceHealth had been using these codes since July 2014 to bill for providers' visits after a hospital discharge, and as of January 2015, practices reported

seeing about 10 patients per week under these codes. PeaceHealth also began to explore the use of the Chronic Care Management codes to bill for medical care coordinator services after the award; they believed that patients with diabetes or hypertension would be eligible for these payments if they received 20 minutes of care coordination per week. However, administrators had concerns that patients might not see the value of paying a $20 Medicare copayment each month and would decline these services. Third, PeaceHealth began looking for community funders and had started communicating with community members about the possibility of starting an endowment. Finally, PeaceHealth is exploring community business risk-share agreements. For example, PeaceHealth had already started talking to the local school system about guaranteeing same-day access for its employees and offering them care coordinator services. The goal would be to expedite access to care for sick employees and help them return to work sooner, generating potential savings for the school system. As of the time of our site visit, however, no firm funding was in place from any source.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

The findings reported by the implementation team—gathered from interviews with program leadership and frontline staff at the two participating sites—provide important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and selected staff might differ from clinicians' views overall. In this section, we use data from the HCIA Primary Care Redesign Clinician Survey (administered in fall 2014, the third year of the HCIA-funded program) to provide additional insight into the implementation process and experience and the contextual factors that might affect implementation effectiveness at PeaceHealth.

In this section, we report on PeaceHealth clinicians' views of their daily work life and practice. First, we focus on the contextual factors that can affect program implementation, including characteristics of the practices' locations, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well their care teams function. We then present data on the alignment of clinicians' views and experiences with the overall goals of the PeaceHealth HCIA-funded innovation, as well as their awareness of and participation in the program and their view of the barriers to and facilitators of successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

Eight clinicians (defined as physicians, nurse practitioners, and physician assistants) work across the two practices participating in the HCIA program. Six clinicians responded to the survey (four physicians and two nurse practitioners), one did not respond, and one was ineligible because he no longer worked at the clinic, resulting in a response rate of 87 percent. Due to the small size of the sample, we are unable to report the number of responses in this section.

Clinicians at both participating sites commonly use health information technology when delivering care. All respondents reported using electronic systems to order lab tests and access their results, prescribe new medications, request prescription and appointment refills, review drug dose or drug interaction warnings, enter clinical notes, and consult with other clinicians. In addition, most respondents reported using electronic patient lists and registries to manage care and track referrals. The survey responses suggest that clinicians widely used the new EHR system, EpicCare, adopted in September 2013, and are consistent with what we heard from frontline staff during our site visit.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall career, opinions about training, and perceptions of their practice environment can all have an effect on the success of program implementation and organizational change. Clinicians were split regarding overall satisfaction with their careers. Some respondents reported they were somewhat satisfied, whereas others reported being either somewhat or very dissatisfied. When asked about their level of burnout, the most common response was "I am definitely burning out and have one or more symptoms of burnout, such as physical and emotional exhaustion." However, most clinicians at least somewhat agreed that their management team was supportive of their concerns, and they felt encouraged by their supervisors to offer suggestions and improvements.

The survey responses also indicated that clinicians' job training and responsibilities do not always align. Although most of the respondents reported that 25 to 49 percent of their work was well-matched to their training, several clinicians reported that most of their work could be done by someone with less training; others said they did not have enough training for at least a quarter of the work they performed. Clinicians were evenly split on whether they had adequate opportunities to develop their professional skills.

Most respondents characterized their workload as unreasonable and, as a result, reported being unable to provide high quality care to all of their patients. Barriers to providing optimal care included lack of time, lack of timely information from other physicians, difficulty obtaining diagnostic tests and treatments, difficulty obtaining specialist referrals, and inadequate reimbursement. Clinicians also said that their patients' inability to pay for care was another barrier to optimal care. Reflecting this concern, PeaceHealth used part of its HCIA funds to hire a financial care coordinator to help patients address their financial needs.

c. Clinicians' perceptions of care team functioning

PeaceHealth used HCIA funds to strengthen its team-based approach to delivering care, including hiring and/or training MOAs and care coordinators. Survey responses reflected this investment: all respondents reported working in care teams and said they were generally satisfied with the support they received. PeaceHealth also hired a social worker and financial counselor to help address patients' psychosocial needs and financial barriers to care. Most respondents reported that care team members relayed relevant information in a timely manner, allowed enough time for questions when communicating with patients, used common terminology when

communicating with one another, verified information they received from one another, and followed a standardized method of sharing information when handing off patients.

d. Alignment with goals of PCR

The survey also asked clinicians to rate the importance of several goals related to PCR, using a scale from extremely important to not important at all. The views of PeaceHealth clinicians generally aligned with the goals of PCR, with most clinicians rating eight of the goals as extremely important and three of the goals as not extremely important (Table II.B.1).

Table II.B.1. Importance of PCR goals

Extremely important	Not extremely important
• Reducing hospital readmissions • Reducing ED visits • Reducing overall health care spending • Improving care coordination for patients with chronic conditions • Improving patients' capacity to manage their own care • Increasing the use of evidence-based practice in clinical care • Improving care continuity in primary care • Increasing the use of EHRs and other health information technology	• Increasing the number of primary care practices functioning as patient-centered medical homes • Improving the capability of health care organizations to provide patient-centered care • Improving the capability of health care organizations to provide team-based care

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: The total number of survey respondents for PeaceHealth was fewer than 11. For this reason, raw data have not been reported.

3. Awareness of program, receipt of training, and perceived effects

a. Awareness of the program and receipt of training

Most respondents said they were familiar with the HCIA program and had received training related to it. On average, clinicians reported receiving slightly fewer than two hours of HCIA training. Training was a large component of the HCIA program at PeaceHealth, but focused mainly on support staff, such as care coordinators and MOAs, rather than clinicians.

b. Perceived effect of program on patients' care

Most respondents reported that the HCIA program had a positive effect on quality of care, ability to respond in a timely way to patients' needs, efficiency, safety, and equity of care for all patients. Although it was not identified as an important goal by respondents, most clinicians thought that the HCIA program also had a positive effect on patient-centeredness.

c. Barriers to and facilitators of program implementation

More than half of responding clinicians identified the following factors as having a positive effect on program implementation: level of program funding, amount of required documentation, availability of personnel, amount of time spent in training or meetings by the program, availability of evidence-based clinical information, availability of community resources to care

for patients with complex conditions, and quality of interpersonal communications with other providers and allied health professionals. Most clinicians said that the quality of interpersonal communications with specialists adversely affected or had no effect on program implementation.

4. Conclusions about clinicians' attitudes and behavior

The survey results suggest that PeaceHealth clinicians are generally satisfied with and supportive of the practice transformations implemented under HCIA funding. They particularly appreciate the shift toward team-based care, including incorporating the new care coordinator and financial coordinator positions and improving the competency levels of the existing medical office assistants. Most clinicians also believe that the HCIA-funded intervention is having a positive effect on patient care and outcomes. However, clinicians reported feeling their workload is unreasonable and experiencing job burnout. Clinicians at PeaceHealth also acknowledged continued challenges coordinating and communicating with specialists, despite investments to improve coordination of care.

C. Impacts on patients' outcomes

1. Introduction

In this part of the report, we draw preliminary conclusions, based on available evidence, about the impacts of PeaceHealth's coordinated care program on Medicare patients' outcomes in three domains: quality-of-care outcomes, service use, and spending. Although PeaceHealth's program serves Medicaid beneficiaries, Medicare fee-for-service (FFS) and managed care beneficiaries, and individuals with private health insurance, due to limitations in available data we have analyzed outcomes only for the Medicare FFS population (including those dually eligible for Medicare and Medicaid). Results might not be generalizable to the full population that PeaceHealth's program serves. In this section, we first describe the methods for estimating impacts (Section II.C.2) and then some characteristics of the PeaceHealth Medicare FFS population at the start of the intervention (Section II.C.3). We next describe the characteristics of the Medicare FFS population at 57 practices in selected towns in Southeast Alaska that serve as a comparison group (Section II.C.4). Finally, in Section II.C.5, we describe the quantitative impact estimates, their agreement with implementation findings, and our conclusions about program impacts in each domain. Our conclusions in this report are preliminary because the analyses do not cover the full time period that we will include in the final impact analysis in future reports.

2. Methods

a. Overview

We estimated program impacts as the difference in outcomes for patients assigned to the two PeaceHealth clinics or practices (we use the terms interchangeably in this section of the report) and outcomes for patients assigned to 57 comparison clinics (or practices), adjusting for any differences between the groups before the coordinated care program began. To focus the analyses, we specified a limited number of primary tests before examining any impact results. Each primary test defined an outcome, population, time period, and direction of expected effects for which we hypothesize impacts if the program is effective. We provided the awardee and CMMI an opportunity to comment on the primary tests, and revised them as appropriate. We

drew conclusions about impacts in each domain based on the results of these primary tests and the consistency of the primary test results with the implementation findings and secondary quantitative tests (robustness and model checks).

b. Treatment group definition

We defined the treatment group separately in each of four baseline quarters before the program began on January 1, 2013 (the baseline period), and in each of eight intervention quarters after the program began (the intervention period). As mentioned earlier, due to limitations in available data, we limited the treatment group for the impact analysis to those PeaceHealth patients enrolled in Medicare FFS, even though this is a subgroup of the total patient population affected by the program. In each quarter of the baseline or intervention period, the treatment group consists of Medicare FFS beneficiaries who (1) were assigned to one of the two treatment clinics on or before the first day of the quarter (see below), (2) had observable outcomes for at least one day in the quarter, and (3) lived in Alaska for at least one day of the quarter. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

Practice assignment. The first step involved attributing beneficiaries to practices using the same decision rule that CMMI uses for the Comprehensive Primary Care Initiative. Specifically, in each baseline and intervention month, we attributed beneficiaries to a practice if the clinic's primary care providers (physicians, nurse practitioners, or physician assistants) provided the plurality of primary care services in the past 24 months. When there was a tie, we attributed the beneficiary to the clinic he or she visited most recently. This attribution method requires identifiers for the providers who worked in the treatment clinics (and when) as well as identifiers for providers in other practices in the region who could compete for patients (when determining which practice provided the plurality of primary care services). PeaceHealth provided data on which providers worked in the two treatment clinics, and SK&A, an outside health care data vendor, supplied identifiers for providers in the other practices. Second, in each period (baseline and intervention), we *assigned* each beneficiary to the first treatment practice he or she was attributed to in the period, and continued to assign him or her to that practice for all quarters in the period.

Using this definition, a beneficiary who has previously been assigned to the treatment group will *remain* a member of the treatment group for the rest of the relevant period (baseline or intervention), as long as he or she is still enrolled in Medicare FFS and living in Alaska by the end of the relevant period. This definition ensures that, during the intervention period, beneficiaries do not exit the treatment group solely because the intervention succeeded in reducing their service use (including visits at treatment clinics). The definition for the baseline period then corresponds to that of the intervention period so that, across the two periods, interpretation of the population changes over time should be comparable.

Definition of high-risk subgroup. Several components of PeaceHealth's coordinated care program focused on providing care to beneficiaries with CHF, diabetes, and/or hypertension. We identified this high-risk subgroup in each quarter by applying Chronic Condition Warehouse algorithms for these conditions to claims in the 12 to 36 months (depending on the condition)

before the start of the baseline or intervention periods. (We did not look at high-risk pregnancies, although this was also a focus of the PeaceHealth program, because pregnancy is rare among Medicare beneficiaries, most of whom are elderly.) As with assignment to the treatment group, a Medicare FFS beneficiary who has previously been identified as having one of these conditions in either period will *remain* a member of this subgroup for the rest of the relevant period (baseline or intervention).

c. Comparison group definition

The comparison group consists of Medicare FFS beneficiaries assigned to 57 comparison practices during each quarter in the baseline and intervention periods. We identified the comparison practices in data we obtained on 239 potential comparison practices in Alaska from SK&A, a health care data vendor. We limited comparison practices to those in geographically isolated parts of Southeast and Southern parts of Alaska, because the PeaceHealth practices are also geographically isolated. We excluded Federally Qualified Health Centers (FQHCs) from the comparison group because neither treatment clinic is an FQHC. These restrictions left us with 57 remaining practices that we used as the comparison group.

We assigned Medicare FFS beneficiaries to the comparison practices in each baseline and intervention quarter using the same rules we used for the intervention group. Further, we defined the subgroup of comparison members with CHF, diabetes, or hypertension in each quarter using the same rules as for the treatment group.

Although we attempted to use propensity-score matching among the 57 potential comparison practices to form a smaller comparison group that would be very similar to each of the two treatment clinics, we were unable to do so. We found that there were no comparison practices that met our minimum criteria of being similar enough to each of the treatment clinics along all of the variables we considered important, including practice size and service use among assigned beneficiaries. After discussions with CMMI, we concluded that the best approach was to have the comparison group include beneficiaries at all 57 practices in the comparison pool, rather than a poorly matched subset. We present the similarity of treatment and comparison groups at baseline in Section II.C.4. In our impact analysis described later (Section II.C.3.e), we controlled for any time-invariant differences in outcomes across practices through practice-level fixed effects.

d. Construction of outcomes and covariates

We used Medicare claims from January 1, 2009, to December 31, 2014, for beneficiaries assigned to the treatment and comparison practices to develop two types of variables: (1) **outcomes,** defined for each person in each baseline or intervention quarter that he or she was a member of the treatment or comparison group; and (2) **covariates** that describe a beneficiary's characteristics at the start of the baseline and intervention periods, and are used in the regression models for estimating impacts to adjust for beneficiaries' characteristics before the period began. We used covariates defined at the start of each period, without updating them each quarter, to avoid controlling in each intervention quarter for previous quarters' program effects, as this

would bias the effect estimates away from detecting true impacts. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. We calculated four quarter-specific outcomes that we grouped into three domains:

1. Domain: Quality-of-care outcomes
   a. Number of inpatient admissions followed by an unplanned readmission within 30 days (number/quarter)
2. Domain: Service use
   b. All-cause inpatient admissions (number/quarter)
   c. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
   d. Total Medicare Part A and B spending ($/month)

All of these outcomes are outcomes that CMMI has specified as core for the evaluations of all HCIA programs. Our definition of the readmission measure, however, differs slightly from CMMI's standard definition. CMMI typically defines readmissions as the proportion of inpatient admissions that end in an unplanned readmission. Instead, we analyze impacts on the *number* of these unplanned readmissions per quarter because this enables us to look at the total impact on readmissions across the treatment group, rather than readmissions contingent on an inpatient admission (because the intervention might affect the number and types of admissions as well).

Covariates. The covariates include (1) whether a beneficiary has each of 10 chronic conditions (including CHF, ischemic heart disease, stroke, chronic obstructive pulmonary disease, hypertension, chronic kidney disease, cancer, diabetes, Alzheimer's and related dementia, or depression). As noted earlier, we identified beneficiaries with these conditions by applying Chronic Condition Warehouse algorithms to claims in the 12 to 36 months (depending on the condition) before the start of the baseline or intervention period; (2) dual Medicare and Medicaid enrollment; (3) Hierarchical Condition Category (HCC) score, which is a continuous score that CMS developed to predict a beneficiary's future Medicare spending; (4) demographics (age, gender, race identified as Native American or Alaska Native versus all other races); and (5) original reason for Medicare entitlement (old age, disability, or end-stage renal disease).

e. Regression model

We used a regression model to implement a difference-in-differences design for estimating impacts. For each quarter-specific outcome, the model estimated the relationship between the outcome and a series of predictor variables, assuming that each of the predictor variables has a linear (additive) relationship with the outcome. The predictor variables included the beneficiary-level covariates (defined in Section II.C.2.d); whether the beneficiary was assigned to a treatment or a comparison practice; an indicator for each practice (which accounts for stable differences

among practices in their outcomes over time); indicators for each post-intervention quarter; and an interaction of a beneficiary's treatment status with each post-intervention quarter. The estimated relationship between the interaction term and outcomes in a given quarter is the impact estimate for that quarter. It measures the average difference between outcomes for beneficiaries assigned to the treatment and comparison practices that quarter, subtracting out any differences between these groups during the four baseline quarters. By providing separate impact estimates for each intervention quarter, the model can identify if the program's impacts changed the longer the clinics are in the program (which is expected to occur). We can also test impacts over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the impact estimates, allowing for statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. The model estimates robust standard errors which account for clustering of outcomes across quarters for the same beneficiary and it includes a dummy variable for each practice (fixed effects) to implicitly account for clustering of outcomes for beneficiaries assigned to the same practice. Appendix 2 provides details on the regression methods.

f. Primary tests

Table II.C.1 shows the primary tests for PeaceHealth, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important (expressed as a percentage of change from the counterfactual—that is, the outcomes that beneficiaries in the treatment group would have had in the absence of the HCIA-funded intervention). The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness. We used information in PeaceHealth's HCIA application to identify the outcomes, time periods, populations, and substantive thresholds for the primary tests (see Appendix 3 for detail and for a description of how we selected each test). We provided both the awardee and CMMI an opportunity to comment on the primary tests.

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** PeaceHealth expected to reduce spending and hospitalizations for all beneficiaries it serves. Among its patients with CHF, diabetes and/or hypertension, it also expected to reduce ED visits. It also expected to improve quality-of-care outcomes, including reducing 30-day unplanned hospital readmissions among patients with CHF.
- **Time period.** PeaceHealth expected program impacts to grow over the first three years before stabilizing in the third year. However, given that the projected impacts were based on the assumption that the intervention would begin soon after the award did, it might be more realistic to expect program effects to be delayed by about a year, given the implementation delay (see Section II.A.2.d). As a result, we conducted the primary tests on outcomes in the second and third intervention years (January 2014 to June 2015, corresponding to the 5th through 10th intervention quarters [I5–I10]), excluding the first intervention year.

Table II.C.1. Specification of the primary tests for PeaceHealth Ketchikan Medical Center

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population	Substantive threshold (impact as percentage of comparison group mean)[c]
Quality-of-care outcomes (1)	30-day unplanned hospital readmission rate (#/person/quarter)	Average over intervention quarters 5 through 10	Medicare FFS beneficiaries with CHF assigned to treatment clinics	-7.5
Service use (2)	All-cause inpatient admissions (#/person/quarter)		Medicare FFS beneficiaries assigned to treatment clinics	-5.0
	Outpatient ED visit rate (#/person/quarter)		Medicare FFS beneficiaries with CHF, diabetes, and/or hypertension assigned to treatment clinics	-15.0
Spending (2)	Medicare Part A and B spending ($/person/month)		Medicare FFS beneficiaries with CHF, diabetes, and/or hypertension assigned to treatment clinics	-12.0
	Medicare Part A and B spending ($/person/month)		Medicare FFS beneficiaries assigned to treatment clinics	-5.0

Note: We used information in PeaceHealth's HCIA application to identify the outcomes, time periods, populations, and substantive thresholds for the primary tests. For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

[a] We adjusted the *p*-values from the primary test results for the multiple comparisons made within each domain, but not across domains.

[b] The regressions we used to estimate impacts controlled for differences in outcomes between the treatment and comparison groups in the baseline period.

[c] We set the substantive threshold to 75 percent of PeaceHealth's expected effect in the second year of the intervention (Section II.C.2.f) for 30-day unplanned hospital readmissions, and for total Medicare Part A and B spending. For the other primary tests, we used the reduction that Peikes et al. (2011) indicated could be feasible among high-risk beneficiaries or general population beneficiaries (as applicable) in a patient-centered medical home program.

CHF = congestive heart failure; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

- **Population.** PeaceHealth's impacts should be concentrated among its high-risk population —specifically those with CHF, diabetes, and/or hypertension—but this population was small compared with the full population served by the HCIA-funded program. In its HCIA proposal, PeaceHealth projected impacts on ED visits that were specific to the chronic condition subpopulations and projected impacts on 30-day unplanned hospital readmissions were specific to those with CHF. However, PeaceHealth also expected reductions in spending among its entire patient population. Because there are trade-offs between analyzing the high-risk subpopulation (for which expected effects are large but the sample size is moderate) and analyzing the entire Medicare FFS population (which is more representative of the program population served but with smaller anticipated effects), we assess both in some of our primary tests, whereas in others we assess impacts only on the subgroup for which PeaceHealth expected to have an impact.
- **Direction (sign) of the impact estimate.** The primary tests are testing for a reduction, relative to the counterfactual, for each of the four outcome measures.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we have specified thresholds for what we call substantive importance. For the primary test populations and outcomes described earlier, the thresholds we chose are the smaller of two values: (1) 75 percent of expected impacts in the second year of the HCIA award, as stated in PeaceHealth's HCIA proposal; and (2) what would be reasonable to expect for this type of intervention based on the literature (Peikes et al. 2011). We use the PeaceHealth estimates for program impacts in the second year of the HCIA award, rather than the third year, because the program was delayed in its implementation. We use a threshold of 75 percent recognizing that a program could be considered successful even if it did not reach its full anticipated effect.

Due to limitations in data availability, we were able to conduct the primary tests in this report only partially. Specifically, we estimated impacts through December 2014, or intervention quarter 8 (I8). As a result, our primary tests in this annual report cover four intervention quarters (I5 through I8). The third annual report should include all intervention quarters, I5 through I10 (January 2014 to June 2015), of the primary tests.

g. Secondary tests

We also conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the non-experimental design or random fluctuations in the data. We will have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. For the secondary tests, we repeated these primary tests, but for outcomes during the first four intervention quarters, the period before PeaceHealth fully implemented its program. Because we expect program impacts to become larger over time, with few or no impacts in the first year of practice participation in the program, the following pattern would be highly consistent with an effective program—few to no measured effects in the first four quarters, growing effects in quarters 5 through 10. In

contrast, if we found very large differences in outcomes (favorable or unfavorable) in the first 12 intervention months, this could suggest a limitation in the comparison group, not true program impacts.

h. Synthesizing evidence to draw conclusions

Within each domain, we drew one of four conclusions about program effectiveness, based on the primary test results and the plausibility of those findings given the implementation evidence. These four possible conclusions are as follows: (1) statistically significant favorable effect (the highest level of evidence), (2) substantively important favorable effect, (3) indeterminate effect, and (4) substantively important unfavorable effect. (We could not conclude that a program had a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program harms.)

We describe our decision rules for each of the four possible conclusions in Appendix 3. In short, we concluded that a program had a statistically significant favorable effect in a domain if (1) at least one primary test result in the domain was favorable and statistically significant, after adjusting the statistical tests to account for multiple tests (if applicable) within a domain; or (2) the average impact estimate across all primary tests in the domain was favorable and statistically significant. In both cases, we also had to determine that the primary test results were plausible given the implementation evidence. We concluded that a program had a substantively important favorable effect if the average impact estimate was substantively important but not statistically significant, and if the result was plausible given the implementation evidence. In contrast, if the average impact estimate was unfavorable (opposite the hypothesized direction), larger than the substantive threshold, and unfavorable effects were plausible given the other evidence, we concluded the program had a substantively important unfavorable effect. Finally, if the tests in a domain did not meet any of these criteria, we concluded that the impact in that domain was indeterminate.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the treatment clinics and beneficiaries at the start of the intervention (January 1, 2013), which can be seen in the second column of Table II.C.2. Each clinic receives a weight equal to the number of beneficiaries assigned to that clinic in the baseline period. For benchmarking purposes, the last column shows the values of relevant variables for the national Medicare population, when available.

Characteristics of the clinics overall. At the start of the intervention, the two treatment clinics, on average, consisted of 9.7 primary care providers (one clinic had 2 whereas the other had 10). Both clinics are owned by the hospital, namely the awardee, PeaceHealth Ketchikan Medical Center. Although both clinics are in remote towns in Alaska, one of the two is in a zip code that the U.S. Census Bureau classifies as urban, and 97 percent of the treatment beneficiaries were assigned to the clinic in the urban zip code. Neither of the clinics is in a borough that is fully designated as having a shortage of health professionals by the federal Health Resources and Services Administration. This may be surprising given the remoteness of

Table II.C.2. Characteristics of treatment and comparison groups when the intervention began (January 1, 2013)

Characteristic of practice	Treatment clinics (N = 2)	Comparison practices (N = 57)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Characteristics of the practices overall					
Practice owned by hospital or health system (%)	100	33	67	0.86	n.a.
Number of PCPs	9.7	4.3	5.44	0.76	n.a.
Characteristics of practices' locations					
Located in an urban zip code (%)	96.7	53.5	43.2	0.65	n.a.
Located in a health professionals shortage area (primary care) (2011) (%)	0.0	40.5	-40.5	-0.69	n.a.
Characteristics of all beneficiaries attributed to practices during the baseline year (January 1, 2012 – December 31, 2012)					
Number of beneficiaries	734	348	386	0.43	n.a.
HCC risk score	1.04	1.00	0.04	0.04	1.0
Unplanned readmissions (#/patient/quarter)	4.3	2.5	1.7	0.31	n.a.
All-cause inpatient admissions (#/1,000 patients/quarter)	70.1	70.1	0.0	0.00	74[c]
Outpatient ED visit rate (#/1,000 patients/quarter)	200.2	141.6	58.6	0.33	105[d]
Medicare Part A and B spending ($/patient/month)	$858	$917	-$58	-0.06	860[e]
Disability as original reason for Medicare entitlement (%)	23.4	22.3	1.1	0.05	16.7[f]
Dually eligible for Medicare and Medicaid (%)	25.4	22.7	2.7	0.12	21.7[g]
Age (years)	70.5	70.7	-0.22	0	71[h]
Female (%)	51.8	52.8	-0.9	-0.02	55.3[f]
Characteristics of high-risk beneficiaries attributed to practices during the baseline year (January 1, 2012 – December 31, 2012)					
Number of high-risk beneficiaries	417	186	230	0.47	n.a.
Unplanned readmissions (#/patient/quarter)	3.9	2.0	1.8	0.42	n.a.
All-cause inpatient admissions (#/1,000 patients/quarter)	89.8	91.5	-1.7	-0.02	n.a.
Outpatient ED visit rate (#/1,000 patients/quarter)	231.3	176.5	54.9	0.23	n.a.
Medicare Part A and B spending ($/patient/month)	1,107	1,150	-42	-0.04	n.a.

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code data merged from the Five-Year American Community Survey ZIP Code Characteristics (2012) and county data merged from the Area Health Resources File (2011).

Notes: Each practice gets a weight equal to the number of beneficiaries assigned to the practice.

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the treatment and comparison groups.

[b] The standardized difference is the difference in means between the treatment and comparison groups divided by the standard deviation of the variable, which is pooled across the treatment and selected comparison groups.

[c] Health Indicators Warehouse (2014b)

[d] Gerhardt et al. (2014).

[e] Boards of Trustees (2013).

[f] Chronic Conditions Data Warehouse (2014, Table A.1).

[g] Health Indicators Warehouse (2014c).

[h] Health Indicators Warehouse (2014a).

*/**/*** Significantly different from zero at the .10/.05/.01 levels, respectively, two-tailed test. No differences were significantly different from zero at the 0.01 level.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; HCC = Hierarchical Condition Category; SD = standard deviation PCP = primary care provider.

n.a. = not applicable.

PeaceHealth primary care clinics, but the designation is largely based on a ratio of population to primary care providers, and given the low population in the two boroughs, they do not meet the criteria for this designation.

Characteristics of the clinics' Medicare FFS beneficiaries. The characteristics of all Medicare FFS beneficiaries assigned to the treatment clinics during the baseline period (January 1, 2012, through December 31, 2012) were similar to the nationwide FFS averages by some measures, but very different by others. The HCC risk score for the treatment group of 1.04 was very close to the national average (1.00). Hospital admission rates (70.1/1,000 people/quarter) and Medicare Part A and B spending ($858/person/month) were close to the national averages, but the outpatient ED visit rate (200.2/1,000 people/quarter) was about twice the national average.

The high-risk beneficiaries in the treatment group had somewhat higher health care utilization and spending during the baseline period than the full treatment group. They had 28 percent more all-cause inpatient admissions, 15 percent more outpatient ED visits, and 29 percent higher Medicare spending. These comparisons are between the high-risk subgroup and the full treatment group; differences would be even larger if we compared the high-risk group to its complement (that is, members of the treatment group who are not a part of the high-risk group).

4. Comparison of the treatment and comparison groups at the start of the intervention

Assessing the similarities and differences between the treatment and comparison groups at the start of the intervention is critical for assessing the quasi-experimental evaluation design and interpreting its results. Similarities increase the credibility of a key assumption underlying difference-in-differences models—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment clinics' beneficiariesnot received the intervention. As discussed in Section II.C.2, we were unable to create a matched comparison group of practices that looked like the treatment clinics on all important, measurable characteristics. As a result, the comparison group consists of all 57 potential comparison practices we identified in geographically isolated areas similar to PeaceHealth in the Southeast and Southern parts of Alaska.

The third column of Table II.C.2 shows weighted mean characteristics at the start of the intervention of the 57 comparison practices. Each practice is weighted by its number of assigned beneficiaries in the baseline period. The comparison practices are smaller on average than the treatment clinics, both by the mean number of primary care providers (4 versus 10) and mean number of beneficiaries (348 versus 734). Although a hospital (PeaceHealth) owns both treatment clinics, a hospital or health system owns only 33 percent of the comparison practices. The comparison practices are less likely to be in zip codes classified as urban by the U.S. Census Bureau and more likely to be located in counties that are identified as health shortage areas. These differences in practice characteristics are all outside of our target of 0.25 standardized differences (the 0.25 target is an industry standard; for example, see Institute of Education Sciences 2014) but in our impact analysis described in Section II.C.3.e, we controlled for these time-invariant differences across practices through practice-level fixed effects.

Despite differences in practice characteristics, beneficiaries attributed to the comparison practices were strikingly similar to those attributed to treatment practices along a number of dimensions. They had similar demographic, health, and eligibility characteristics—age, gender, HCC risk scores, percentage with dual Medicare and Medicaid coverage, and percentage with disability as reason for original Medicare entitlement. The groups also had very similar mean rates of admissions and Medicare Part A and B spending. However, the comparison group had substantially lower unplanned readmission and outpatient ED visit rates.

As noted previously, in our impact analysis we controlled for differences across practices through practice-level fixed effects and through beneficiary-level covariates. Nevertheless, because some of the baseline characteristics of the treatment and comparison groups do differ, results should be interpreted with caution.

5. Intervention impacts

In this section, we first present sample sizes and mean outcomes, by cohort and quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates; however, the differences in mean outcomes are not impact estimates by themselves. Next, we present the results of the primary tests (which are regression-adjusted), by domain. Then, we present the secondary tests results and assess whether the primary test results are plausible given the secondary tests. Next, we assess whether primary test results are plausible given the implementation evidence. We end with preliminary conclusions about program impacts in each domain.

a. Sample sizes

We present sample sizes for all assigned beneficiaries and two subpopulations included in our primary tests: (1) high-risk beneficiaries, defined as those with CHF, diabetes or hypertension; and (2) beneficiaries with CHF only. Over the four baseline quarters the number of assigned beneficiaries increased from 714 to 886 in the treatment group and from 9,242 to 10,860 in the comparison group (Table II.C.3). Similarly over the eight intervention quarters, the number of attributed beneficiaries increased from 846 to 1,046 in the treatment group and from 9,657 to 11,404 in the comparison group. The number of high-risk beneficiaries in the baseline period ranged from 467 to 522 among the treatment group and from 5,996 to 6,454 among the comparison group. In the intervention quarters, these samples ranged from 550 to 569 for the treatment group and 6,319 to 6,522 for the comparison group. The sample size for attributed beneficiaries with CHF in the baseline quarters ranged from 47 to 60 among the treatment group and 990 and 1,018 among the comparison group. During the intervention quarters, these samples ranged from 66 to 74 for the treatment group and 925 to 1,066 for the comparison group.

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

Quality-of-care outcomes. The number of 30-day unplanned readmissions, which was estimated only among beneficiaries with CHF, ranged from 0 to 38.5 per 1,000 beneficiaries per quarter for the treatment group and 27.5 to 40.4 for the comparison group during the baseline quarters. The readmission rate ranged from 0 to 81.1 for the treatment group and 27.0 to 40.3 for the comparison group in the eight intervention quarters.

Table II.C.3. Sample sizes and unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and comparison groups for PeaceHealth Ketchikan Medical Center, by quarter

	Number of Medicare FFS beneficiaries (practices)						30-day unplanned hospital readmission rate for CHF (#/1,000/month)			All-cause inpatient admissions for all FFS beneficiaries(#/1,000/quarter)			Outpatient ED visit rate for high risk (#/1,000/quarter)			Medicare Part A and B spending for all FFS beneficiaries ($/month)			Medicare Part A and B spending for high risk ($/month)		
	Total		High risk		CHF																
Q	T	C	T	C	T	C	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (January 1, 2012 – December 31, 2012)																					
B1	714 (2)	9,242 (57)	467	5,966	47	990	0.0	40.4	-40.4 (-100.0%)	77.0	77.4	-0.3 (-0.4%)	265.5	156.9	108.6 (69.2%)	$866	$943	$-77 (-8.1%)	$926	$1,111	$-185 (-16.7%)
B2	774 (2)	9,849 (57)	489	6,175	52	1,000	38.5	29.0	9.5 (32.6%)	68.5	67.7	0.8 (1.1%)	204.5	154.5	50.0 (32.4%)	$799	$884	$-84 (-9.6%)	$1,031	$1,054	$-23 (-2.2%)
B3	843 (2)	10,421 (57)	516	6,328	59	1,009	16.9	32.7	-15.8 (-48.2%)	70.0	62.9	7.1 (11.4%)	269.4	167.5	101.9 (60.8%)	$811	$861	$-50 (-5.8%)	$1,023	$1,003	$20 (2.0%)
B4	886 (2)	10,860 (57)	522	6,454	60	1,018	33.3	27.5	5.8 (21.2%)	70.0	66.9	3.1 (4.7%)	182.0	165.9	16.0 (9.7%)	$987	$910	$77 (8.4%)	$1,319	$1,101	$218 (19.8%)
Intervention period (January 1, 2013 – December 31, 2014)																					
I1	846 (2)	9,657 (57)	550	6,319	71	1,065	28.2	34.7	-6.6 (-18.9%)	80.4	71.6	8.8 (12.3%)	212.7	164.9	47.8 (29.0%)	$1,019	$942	$77 (8.2%)	$1,303	$1,164	$139 (12.0%)
I2	894 (2)	10,194 (57)	557	6,473	74	1,066	67.6	40.3	27.2 (67.5%)	63.8	67.9	-4.1 (-6.1%)	213.6	160.2	53.4 (33.4%)	$748	$918	$-170 (-18.5%)	$901	$1,088	$-187 (-17.2%)
I3	936 (2)	10,566 (57)	567	6,512	74	1,058	81.1	38.8	42.3 (109.2%)	54.5	64.7	-10.2 (-15.8%)	227.5	156.0	71.5 (45.8%)	$819	$966	$-147 (-15.2%)	$948	$1,156	$-208 (-18.0%)
I4	964 (2)	10,852 (57)	566	6,522	71	1,028	14.1	31.1	-17.0 (-54.8%)	59.1	69.6	-10.4 (-15.0%)	199.6	154.9	44.8 (28.9%)	$778	$952	$-174 (-18.3%)	$846	$1,143	$-296 (-25.9%)
I5	994 (2)	11,026 (57)	569	6,482	72	997	13.9	25.1	-11.2 (-44.6%)	57.3	69.7	-12.4 (-17.8%)	237.3	152.3	85.0 (55.8%)	$894	$958	$-65 (-6.7%)	$1,082	$1,158	$-76 (-6.5%)
I6	999 (2)	11,212 (57)	560	6,440	69	966	14.5	32.1	-17.6 (-54.8%)	71.1	64.8	6.2 (9.6%)	283.9	159.1	124.8 (78.5%)	$1,169	$992	$178 (17.9%)	$1,465	$1,182	$283 (23.9%)
I7	1,028 (2)	11,419 (57)	560	6,386	68	950	14.7	29.5	-14.8 (-50.1%)	71.0	60.3	10.7 (17.7%)	212.5	166.2	46.3 (27.8%)	$1,036	$984	$52 (5.3%)	$1,250	$1,193	$57 (4.8%)
I8	1,046 (2)	11,404 (57)	556	6,323	66	925	0.0	27.0	-27.0 (-100.0%)	54.5	67.9	-13.4 (-19.7%)	199.6	156.2	43.5 (27.8%)	$810	$916	$-107 (-11.7%)	$1,018	$1,097	$-79 (-7.2%)

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Table II.C.3 *(continued)*

Note: The baseline quarters are measured relative to the start of the baseline period on January 1, 2012. For example, the first baseline quarter (B1) ran from January 1, 2012, to March 31, 2012. The intervention quarters are measured relative to the start of the intervention period on January 1, 2013. For example, the first intervention quarter (I1) runs from January 1, 2013, to March 31, 2013. In each period (baseline or intervention), the treatment group each quarter includes all beneficiaries assigned to a treatment practice by the start of the quarter and who met other sample criteria—that is, they were enrolled in FFS Medicare, were living in Alaska, and were observable. In each period, the comparison group each quarter includes all beneficiaries assigned to a comparison practice by the start of the quarter and who met the other sample criteria. See text for details.

B = baseline; C = comparison; Diff = difference; ED = emergency department; FFS = fee-for-service; I = intervention; T = treatment; Q = quarter.

NA = not available.

n.a. = not applicable

Service use. The number of all-cause inpatient admissions, which was estimated among all assigned beneficiaries, ranged from 68.5 to 77.0 per 1,000 beneficiaries per quarter among the treatment beneficiaries and 62.9 to 77.4 among the comparison beneficiaries during the baseline quarters. The rate ranged from 54.5 to 80.4 all-cause hospital admissions among treatment beneficiaries in the eight intervention quarters. The mean rates were lower among the treatment group than the comparison group in five of the eight intervention quarters. The differences ranged from -19.7 to 17.7 percent across the eight intervention quarters, with no clear pattern over time.

The number of ED visits, which was estimated only among high-risk beneficiaries, ranged from 182.0 to 269.4 per 1,000 beneficiaries per quarter for the treatment group during the baseline quarters. This rate was consistently higher than the number of ED visits among the comparison group during the baseline quarters, which ranged from 154.5 to 167.5. During the intervention quarters, ED visits ranged from 199.6 to 283.9 among treatment beneficiaries. The mean during the intervention quarters was also consistently higher among the treatment group than the comparison group, ranging from 27 to 46 percent higher.

Spending. We estimated total Medicare Part A and B spending both for all assigned beneficiaries and the subgroup of high-risk beneficiaries. Among all assigned beneficiaries, spending per person per month ranged from $799 to $987 among treatment beneficiaries and $861 to $943 among comparison beneficiaries in the baseline quarters. During the eight intervention quarters, the treatment group mean ranged from $748 to $1,169. Mean spending was lower among the treatment group than the comparison group in five of the eight intervention quarters. The difference ranged from -15.5 to 17.9 percent, but with no clear pattern over time. Among high-risk beneficiaries, spending in the baseline quarters ranged from $926 to $1,319 per person per month among treatment beneficiaries and $1,003 to $1,111 among comparison beneficiaries. During the eight intervention quarters, the treatment group mean ranged from $846 to $1,465. Mean spending was lower for the treatment group than the comparison group in five of the eight intervention quarters. The difference ranged from -25.9 to 23.9 percent, with no clear pattern over time.

c. Results for primary tests, by domain

Overview. The primary test results reflect the average impact estimate in I5 through I10. For this report, we had data available for only the fifth through eighth quarters of this period. Thus, the primary tests in this report reflect impacts over only four of the six relevant intervention quarters.

Results of the primary tests differed by domain (Table II.C.4). Tests in the quality-of-care outcomes domain indicate substantively important but not statistically significant favorable effects. Tests in the service use and spending domains were indeterminate. Table II.C.4 also shows that in general, the tests had poor statistical power to detect effects of substantive importance. As described earlier, these results are preliminary because the analyses do not yet cover the full time period that we will include in the final impact analysis in future reports.

Table II.C.4. Results of primary tests for PeaceHealth Ketchikan Medical Center

Primary test definition					Statistical power to detect an effect that is[b]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual)[a]	Size of the substantive threshold	Twice the size of the substantive threshold[c]	Treatment group mean	Regression-adjusted difference between treatment group mean and counterfactual (standard error)[a]	Percentage difference[d]	*p*-value
Quality-of-care outcomes (1)	30-day unplanned readmissions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 5 through 8	All observable Medicare FFS beneficiaries with CHF assigned to treatment clinics	-7.5%	12.2	14.8	10.8	-19.7 (19.2)	-64.6	0.15[f]
Service use (2)	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)		All observable Medicare FFS beneficiaries assigned to treatment clinics	-5.0%	20.2	35.0	63.5	-8.7 (8.1)	-12.0	0.23[f]
	Outpatient ED rate (#/1,000 beneficiaries/quarter)		All observable Medicare FFS beneficiaries with CHF, diabetes, and/or hypertension assigned to treatment clinics	-15.0%	54.9	93.7	233.3	13.3 (23.5)	6.0	0.60[f]
	Combined (%)			-10.0%	43.7	83.3	n.a.	n.a.	-3.0	0.37[g]
Spending (2)	Medicare Part A and B spending ($/beneficiary/month)		All observable Medicare FFS beneficiaries assigned to treatment clinics	-5.0%	21.2	37.7	977	7 (100)	0.7	0.50[f]
	Medicare Part A and B spending ($/beneficiary/month)		All observable Medicare FFS beneficiaries with CHF, diabetes, and/or hypertension assigned to treatment clinics	-12.0%	36.6	72.4	1,204	-46 (160)	-3.6	0.45[f]
	Combined			-8.5%	30.0	59.2	n.a.	n.a.	-1.5	0.45[g]

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Note: The results for each outcome are based on a difference-in-differences regression model, as described in the text. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

Table II.C.4 *(continued)*

[a] The counterfactual is the outcome the treatment group beneficiaries would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[b] The power calculation is based on actual standard errors from analysis. For example, in the first row, a 7.5 percent effect on 30-day unplanned readmissions (from the counterfactual of 10.8 + 19.7 = 30.5) would be a change of 2.3 readmissions. Given the standard error of 19.2 from the regression model, we would be able to detect a statistically significant result 12.2 percent of the time if the impact was truly 2.3 readmissions, assuming a one-sided statistical test at the $p = 0.10$ significance level.

[c] We show statistical power to detect a very large effect (twice the size of the substantive threshold) because this provides additional information about the likelihood that we will find effects if the program is indeed effective. If power to detect effects is less than 75 percent even for a very large effect, then the evaluation is extremely poorly powered for that outcome.

[d] Percentage difference is calculated as the regression-adjusted difference-in-differences estimate, divided by the estimate of the counterfactual.

[e] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test.

[f] We adjusted the *p*-values from the primary test results for the multiple (two) comparisons made within the domain.

[g] This *p*-value tests the null hypothesis that the difference-in-differences estimates across the two outcomes in the domain, each expressed as percentage change from the estimated counterfactual, is greater than or equal to zero (a one-sided test).

CHF = congestive heart failure; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

n.a. = not applicable.

Quality-of-care outcomes. The 30-day unplanned readmission rate, which was estimated among beneficiaries with CHF, averaged 10.8 per 1,000 beneficiaries per quarter during the primary test period for the treatment group, which was estimated to be 20.0 readmissions (65 percent) fewer than the counterfactual. (Our estimate of the counterfactual is the treatment group mean minus the difference-in-differences estimate.) This difference is much larger than the substantive threshold of 15 percent but not statistically significant ($p = 0.15$).

Service use. All-cause inpatient admissions per 1,000 beneficiaries, estimated among all assigned beneficiaries, averaged 63.5 per quarter during the primary test period among the treatment group. This was estimated to be 8.7 admissions (12 percent) fewer than the counterfactual. The rate of outpatient ED visits, estimated among high-risk beneficiaries, averaged 233.3 visits per 1,000 beneficiaries over the primary test period among the treatment group, 13.3 visits (6 percent) *higher* than the counterfactual. However, neither difference was statistically significant after adjusting for multiple statistical tests in the domain. The mean impact over the two service use outcomes was -3 percent. This did not meet our threshold for substantive importance (-10 percent); nor was it statistically significant ($p = 0.38$). Our statistical power to detect an effect the size of the substantive threshold was poor (47.3 percent). This means that, if the program had succeeded in reducing service use by 10 percent, our analysis would have had only a 47.3 percent probability of detecting that effect (using a one-tailed test and a $p < 0.10$ threshold).

Spending. Medicare Part A and B spending for all beneficiaries assigned to the treatment group averaged $977 per beneficiary per month over the primary test period. This was estimated to be $7 or 0.7 percent higher than the counterfactual, though this difference was not statistically significant. Spending for high-risk beneficiaries during the primary test period averaged $1,204, which was estimated to be $46 or 3.6 percent lower than the counterfactual. The mean percentage difference across the two spending outcomes was -1.5 percent and not statistically significant ($p = 0.45$). As with service use, statistical power to detect an effect the size of the substantive threshold was poor (30 percent).

d. Results for secondary tests

For the secondary tests (Table II.C.5), we repeated the primary tests described previously, but on I1 through I4, the four intervention quarters before PeaceHealth fully implemented its program.

For 30-day unplanned readmissions, the secondary tests show a substantively large difference between the treatment group and the counterfactual (23 percent). However, this difference reflects a *higher* number of readmissions among the treatment group during this period. The large change in the difference from secondary test period (I1 through I4) to the primary test period (I5 through I8) might reflect noise in the data due to small sample sizes in both periods.

In the service use domain, the secondary test results show a lower rate of all-cause inpatient admissions among the treatment group relative to the counterfactual, but not above the

Table II.C.5. Results of secondary tests for PeaceHealth Ketchikan Medical Center

Secondary test definition				Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between treatment group mean and counterfactual (standard error)[a]	Percentage difference [b]	*p*-value
Quality-of-care outcomes (1)	30-day unplanned readmissions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 1 through 4	All observable Medicare FFS beneficiaries with CHF assigned to treatment clinics	47.7	8.9 (25.1)	23.0	0.64[c]
Service use (2)	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)		All observable Medicare FFS beneficiaries assigned to treatment clinics	64.4	-10.1 (8.4)	-13.6	0.11[c]
	Outpatient ED rate (#/1,000 beneficiaries/quarter)		All observable Medicare FFS beneficiaries with CHF, diabetes, and/or hypertension assigned to treatment clinics	213.4	-6.6 (21.8)	-3.0	0.38[c]
Spending (2)	Medicare Part A and B spending ($/beneficiary/month)		All observable Medicare FFS beneficiaries assigned to treatment clinics	841	-110* (83)	-11.6	0.09[c]
	Medicare Part A and B spending ($/beneficiary/month)		All observable Medicare FFS beneficiaries with CHF, diabetes, and/or hypertension assigned to treatment clinics	1,000	-218** (120)	-17.9	0.03[c]

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Notes: The analyses in Table II.C.5 were conducted in the same way as the analyses in Table II.C.4. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

The *p*-values from the secondary test results were *not* adjusted for multiple comparisons within each domain or across domains.

[a] The counterfactual is the outcome the treatment group beneficiaries would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[b] Percentage difference is calculated as the regression-adjusted difference-in-differences estimate, divided by the estimate of the counterfactual.

[c] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test.

CHF = congestive heart failure; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

n.a. = not applicable.

CHF = congestive heart failure; CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service.

substantive threshold and only borderline significant ($p = 0.11$). For ED visits, the secondary tests show no large difference between the treatment group and the counterfactual.

In the spending domain, the secondary tests show significantly lower spending among the treatment group than the counterfactual.

The secondary tests for spending suggest there could be some unobservable differences between treatment and comparison groups that might partially drive results. This finding is not surprising, given that we were unable to construct a comparison group that was well balanced on all important baseline characteristics; as a result, there are some differences between the two groups on important variables, and these could be correlated with other, important and unobserved variables that affect outcomes. However, the secondary tests for the service use domain suggest there were no important differences between the two groups on these outcomes in the first year of the intervention, and the substantive difference for 30-day readmissions was likely driven by small sample sizes. Overall, the secondary tests remind us to interpret the results of our primary tests cautiously, given potential unobserved differences between the treatment and comparison practices.

e. Consistency of quantitative estimates with implementation findings

Our quantitative estimates are plausible based on findings from implementation analyses. Although none of our quantitative estimates are statistically significant, substantively favorable impacts on quality of care for CHF patients is consistent with implementation findings that CHF patients received the most intensive intervention services.

f. Conclusions about program impacts, by domain

Based on all evidence currently available, we have drawn the following preliminary conclusions about program impacts in each domain (as summarized in Table II.C.6).

- **Quality-of-care outcomes**. The magnitude of the estimates are consistent with a substantively important favorable effect of the program on this domain among CHF patients. Although the estimated mean effect was larger in magnitude than our threshold for substantive importance, it was not statistically significant. The lack of statistical significance might be due to the limited statistical power given the small sample sizes. Unobserved differences between treatment and comparison groups, as described earlier, could also drive the results.
- **Service use.** The program had an indeterminate effect on service use. The estimated effects on the two outcomes in the domain (admissions among all assigned beneficiaries and ED visits among high-risk beneficiaries) were not statistically significant, and the mean estimated effect did not meet our threshold for substantive importance. The indeterminate effect in the service use domain could mean one of two things. It is possible the program did not have an effect on service use. Alternatively, the program might have had some small effect—and possibly one that exceeded the substantive threshold—but we were unable to detect it given the small sample sizes available in the claims data used for this report.

Table II.C.6. Preliminary conclusions about the impacts of PeaceHealth Ketchikan Medical Center, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion		
		Primary test result(s) that supported conclusion	Primary test result plausible given secondary tests?[a]	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	Substantively important (but not statistically significant) favorable effect	• The estimated effect of the one test in the domain was substantively important but not statistically significant	Yes	Yes
Service use	Indeterminate effect	• No individual tests in the domain were statistically significant • The mean effect across all tests in the domain was neither substantively important nor statistically significant	Yes	Yes
Spending	Indeterminate effect	• No individual tests in the domain were statistically significant • The mean effect across all tests in the domain was neither substantively important nor statistically significant	Yes	Yes

Sources: Tables II.C.4 and II.C.5.

[a] Although the results of the primary tests are plausible given the results of the secondary tests, the results of the secondary tests suggest that unobserved differences between the treatment and comparison groups could have affected outcomes.

Finally, the indeterminate effect could be driven by unobservable differences between treatment and comparison groups, as described in the secondary test section.

- **Spending.** The program had an indeterminate effect on this domain. The mean effect across the two outcomes (Medicare Part A and B spending among all attributed beneficiaries and Medicare Part A and B spending among high-risk beneficiaries) was neither substantively large nor statistically significant. As with the other domains, this could reflect that the program did not have an effect on spending, or that it had an effect—and possibly one that exceeded the substantive threshold—but, due to the statistical uncertainty in the estimates, we were unable to detect it. However, the impact estimates for spending from the secondary test period, before the PeaceHealth program was fully implemented, suggest that differences observed in Medicare spending between treatment and comparison groups could be due to something other than the HCIA-funded program

As mentioned previously, these conclusions are preliminary because the analyses do not yet cover the full time period that we will include in the final impact analysis in future reports.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

PeaceHealth received HCIA funding to implement a multidimensional coordinated care program. The goals of the program were to (1) improve access to primary care by hiring staff and increasing after-hours care; (2) increase support to and improve outcomes for high-risk patients (particularly those with diabetes, CHF, and, later, hypertension and high-risk pregnancies) by hiring care coordinators to manage chronic conditions and link patients to community resources; and (3) strengthen primary care teams by enhancing the skills of medical assistants and implementing routine scrub-and-huddle procedures. The program involved four interrelated components: (1) transitional care services for all patients discharged from the PeaceHealth Ketchikan Medical Center and intensive transitional care services for patients with CHF, (2) short-term care management for patients with a temporary medical or social hurdle, (3) long-term case management for patients requiring assistance to effectively manage their chronic conditions, and (4) population health management including redefining the scrub-and-huddle process and outreach to paneled patients to improve preventive care. Providing frontline users flexibility to tailor services to the needs of their patients, dedicating sufficient resources to support the program, and aligning program goals with the culture of the organization facilitated program implementation. An initial lack of engagement among referring clinicians, lack of clarity in roles and protocols for new positions, unexpected adoption of a new EHR system, and difficulty recruiting staff in an isolated region of Alaska hindered implementation. The HCIA-Primary Care Redesign Clinician Survey findings suggest that PeaceHealth clinicians grew to become generally satisfied with and supportive of the practice transformations, especially the delegation of care coordination and care management responsibilities among staff with less clinical training and the shift toward team-based care.

The impact evaluation found a substantively important favorable effect on quality-of-care outcomes, but no measurable effects on service use or Medicare Part A and B spending during the first 12 months of the primary test period (January 2014 through December 2014). The lack of statistically significant effects might be because the program did not have any effects or that it did but, due to poor statistical power, our tests failed to detect them. The program could have measurable impacts in one or more of three domains when the evaluation is extended to cover the full primary test period (January 2014 through June 2015).

Our next steps for this evaluation are to (1) monitor PeaceHealth's program implementation reports through June 30, 2015, and plans for sustaining the program beyond the funding period; (2) evaluate trainees' and clinicians' attitudes and experiences with the program in the third year of the award through administered surveys; (3) extend the impact evaluation to include the full period of program operations; and (4) use the implementation findings to help interpret the impact results.

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

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table A.1 Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: CMS, 2014. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: Health Indicators Warehouse, National Center for Health Statistics, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: Health Indicators Warehouse, National Center for Health Statistics, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Health Indicators Warehouse. "Medicare Beneficiaries Eligible for Medicaid (percent)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014c. Available at http://www.healthindicators.gov/Indicators/Medicare-beneficiaries-eligible-for-Medicaid-percent_317/Profile/ClassicData. Accessed August 4, 2015.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Research Institute at Nationwide Children's Hospital

March 2016

Joseph Zickafoose
Brenda Natzke
Boyd Gilman
Catherine DesRoches
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

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

RESEARCH INSTITUTE AT NATIONWIDE CHILDREN'S HOSPITAL

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by Nationwide Children's Hospital (NCH) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the NCH program. Section II presents a summary of the evaluation findings about the effectiveness of program implementation. A primary care clinician survey was not fielded for this awardee, so a description of the attitudes and behaviors of the clinicians affected by the program is not available for this awardee. The third evaluation component—estimating the impact of the program on patient outcomes—also is not available for this awardee due to delays in obtaining Medicaid data. In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF NCH

NCH received a three-year, $13.2 million HCIA in partnership with Akron Children's Hospital (ACH) and Partners for Kids (PFK) to improve care and health and lower costs for children enrolled in Medicaid managed care. Table I.1 summarizes key features of the program. NCH and its partners aimed to replicate an existing Medicaid accountable care organization (ACO) for children, improve care for children with complex chronic conditions and behavioral health care needs served by NCH and ACH, and reduce the rates of preterm births and related neonatal hospital care in Summit County, Ohio. Program goals (core outcomes) included (1) reducing per-member per-month costs for children enrolled in Medicaid managed care due to disability by 2 percent, (2) reducing per-member per-month costs for other children enrolled in Medicaid managed care by 1 percent, (3) reducing hospital inpatient days for children with feeding tubes by 10 percent, (4) increasing the number of tube-fed children with healthy weights by 10 percent, (5) providing proactive care coordination for 85 percent of tube-fed children with a neurological diagnosis, (6) reducing behavioral health-related hospital readmissions by 30 percent, (7) reducing post-discharge impairment for hospitalized behavioral health patients by at least 15 percent, (8) increasing the rate of outpatient mental health follow-up appointments within 30 days following discharge to at least 85 percent, (9) increasing progesterone use in pregnant mothers in Summit County with previous preterm births by 10 percent, (10) reducing the preterm birth rate in Summit County by 20 percent, and (11) reducing neonatal intensive care unit days at ACH by 10 percent. NCH received a no-cost extension to continue program activities through December 2015.

Table I.1. Summary of NCH PCR program

Awardee's name	Research Institute at Nationwide Children's Hospital
Award amount	$13,160,092
Implementation date	November 2012
Award end date	December 2015
Program description	1. Replicate a Medicaid accountable care organization (ACO) model in northeast Ohio 2. Enroll children eligible for Medicaid based on disability into an existing Medicaid ACO 3. Improve care and lower costs for children with behavioral health care needs 4. Improve care and lower costs for children with complex chronic conditions 5. Lower the rate and reduce the cost of premature births in Summit County, Ohio
Innovation components	Care coordination, care management, care transitions, health IT, patient-centered care, integrated team care, payment reform, workflow redesign, home care
Intervention focus	Individual
Workforce development	Create new positions (complex care and behavioral health care coordinators, parent peer partners)
Target population	Medicaid-enrolled children, children with chronic conditions, children with behavioral and mental health disorders, mothers with previous premature births
Program setting	Provider (hospital-based)
Market area	Regional (46 of 88 counties in Ohio)
Market location	Urban and rural
Core outcomes	• 2 percent reduction in cost of care for children enrolled in Medicaid based on disability • 1 percent reduction in cost of care for other Medicaid-enrolled children • 10 percent reduction in hospital days for children with feeding tubes • 10 percent increase in the percentage of tube-fed children with healthy weights • 85 percent increase in enrollment in proactive care coordination for tube-fed children with a neurological diagnosis • 30 percent reduction in behavioral health hospital readmissions • 15 percent decrease in impairment of behavioral health care patients 30 days after discharge • 85 percent rate of follow up with mental health provider within 30 days after discharge • 10 percent increase in progesterone use by pregnant mothers with prior preterm births • 20 percent reduction in preterm birth rate • 10 percent reduction in neonatal intensive care unit days

Source: Review of NCH program reports, March 2015.

Note: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health IT systems, and undertaking other operational activities. The award end date was extended from June 30, 2015, to December 31, 2015, through a partial no-cost extension awarded by CMMI.

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how it has been adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss findings related to

program sustainability and scalability. We based our evaluation of NCH's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collecting during site visits conducted in March 2014 and March 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

NCH and ACH are freestanding pediatric tertiary care hospitals located in Columbus and Akron, Ohio, respectively. In addition to inpatient care, both offer outpatient primary, specialty, and emergency care services throughout their regions. PFK is a physician hospital organization formed by a partnership between NCH and independent providers throughout a 34-county region in central and southeast Ohio. PFK began operations in 1994 and has evolved into a Medicaid ACO model that covered care for approximately 300,000 children in Medicaid managed care at the time the HCIA began.

Table II.A.1 presents key details about the primary program components that enrolled patients and their families. First, NCH and ACH both undertook interventions to improve care for children with behavioral health care needs, including formal peer-to-peer support for parents of children with behavioral health care needs (parent peer partners) and behavioral health care coordination. The parent peer partner intervention aimed to improve the patient- and family-centeredness of care by hiring parents with significant experience with the behavioral health care system through their own children to work with caregivers of children admitted to behavioral health care units and providers. Both NCH and ACH also created roles for behavioral health care management and coordination, but the duration and intensity of the intervention differed between the organizations. In addition to these core interventions, NCH used award funding to pilot the use of technology to enhance behavioral health care, including the use of an online therapy platform for recently hospitalized patients and patients in outpatient treatment programs, and development of a dialectic behavior therapy mobile application for families already involved in a direct clinical program as a way to improve access to resources and management plans. In spring 2015, NCH planned to expand the use of online therapy for recently hospitalized patients.

Second, NCH and ACH implemented several interventions to improve care for children with complex care needs, including children who have neurological conditions and a feeding tube and children with tracheostomies. These interventions included (1) multidisciplinary care coordination teams; (2) the development of standardized hospital-based care protocols for children with technology dependence, such as feeding tubes or tracheostomy; and (3) development of resources to support family education and self-management of their child's needs. The primary adaptations of this intervention were efforts to make workflows more efficient—for example, by providing input into changes to the electronic health record (EHR) care plan format to make it more user friendly (adding boxes and pull-down menus instead of relying on free text), which decreased time spent on initial health assessments and creation of a care plan. Both NCH and ACH worked with multiple specialty departments providing care for

Table II.A.1. Key details about program design and adaptation

	Program component			
	Behavioral health parent peer partners (patient-centered care)	**Behavioral health care coordination and management**	**Complex care coordination and care process standardization**	**Prematurity prevention**
Target population	Caregivers of children with behavioral health needs admitted to crisis intervention, inpatient behavioral health care, and inpatient medical units	Caregivers of children with high behavioral health care needs, as identified by an inpatient admission or a provider referral	Children with complex care needs (initially a feeding tube and a neurological condition; in 2014 added tracheostomies) and their caregivers	Summit County, Ohio: women with a prior history of a premature delivery; obstetrics providers
Identification strategy	Parent peer partners review of admissions to behavioral health care units, provider referrals	Program staff identification of discharges from an NCH or ACH behavioral health care unit, Medicaid MCO referral, provider referral	Program analysts analysis of internal administrative and billing data, provider referrals	Neonatal intensive care unit staff identify all mothers of infants born prematurely; component coordinator identifies all obstetrics providers in the county
Recruitment/ enrollment strategy	Caregivers contacted directly by parent peer partners on day of admission or referral	NCH – telephone call followed by an initial in-person meeting ACH – telephone call following discharge	Telephone call followed by an initial in-person meeting	Neonatal intensive care unit nursing staff provide education to mothers; component coordinator reaches out to obstetrics providers
Service delivery protocol	Employed parent peer partners; daily or more frequent in-person interactions with child's caregivers beginning within 24 hours of admission; daily participation in behavioral health team rounds	NCH – intake needs assessment; in-person meetings with caregivers every 3 months at minimum; interval telephone, text, and email contacts ACH – initial call 1 or 2 days after discharge, follow-up calls a few days before and after first outpatient follow-up visit, other contacts as needed	Intake needs assessment, including medical, nursing, home care, nutritional, and social needs; in-person meetings with caregivers every 3 months at minimum; interval telephone, text, and email contacts; standardization of inpatient care protocols; development of caregiver self-management tools available online	Education on the effectiveness of progesterone therapy for preventing repeat premature deliveries to mothers and financial support for treatment for uninsured mothers; education on overcoming insurance and other logistical barriers to delivery of progesterone to obstetrics providers

Table II.A.1 *(continued)*

	Program component			
	Behavioral health parent peer partners (patient-centered care)	**Behavioral health care coordination and management**	**Complex care coordination and care process standardization**	**Prematurity prevention**
Adaptations	Additional training in motivational interviewing, adjustment of staffing and personnel policies to meet needs of parent peer partners whose own children could have crises, spread of service to additional hospital units; initially, all children with behavioral health needs were included, but later targeting of the component was refined based on level of need and insurance type (Medicaid)	No major adaptations aside from ongoing adjustments to documentation procedures	ACH developed a tiering system to identify children who needed more or less in-person contact	ACH began additional outreach to pediatric primary care providers who care for a significant number of premature infants to provide educational materials for providers and mothers of the infants

Source: Interviews from second site visit, March 2015; document review, March 2015.
ACH = Akron Children's Hospital; MCO = managed care organization; NCH = Nationwide Children's Hospital.

the target population children to standardize care processes, such as selection of initial feeding tube equipment and documentation of feeding plans. Both teams modified their EHRs to support the standardization of care and coordinate information across providers.

Third, ACH built on prior work done by NCH and PFK to reduce the incidence of premature births and related neonatal hospital care. ACH sought to improve communitywide adoption of an evidence-based intervention to reduce the risk of premature delivery, progesterone therapy for pregnant women with a prior history of a premature birth. ACH worked to identify at-risk populations of women, primarily in three neonatal intensive care units in Summit County, Ohio, to identify mothers with spontaneous preterm birth admissions and provide educational materials about the need for progesterone therapy in future pregnancies. The program also sought to engage and educate obstetrics providers, offering them educational materials for their patients and guidance to providers about navigating insurance and other barriers to delivery of the treatment.

In addition to these four interventions focused on patients and families, NCH's HCIA had two administrative components. First, NCH and PFK provided guidance to ACH in developing its own ACO administrative and operational infrastructure and in negotiating contracts with Medicaid managed care organizations (MCOs). Second, PFK prepared for enrollment of children

eligible for Medicaid based on disability into the ACO. These children were previously excluded from Medicaid managed care, but state policy changed. HCIA funding primarily supported PFK to develop the data analytics staff and infrastructure to begin to assess the health care utilization and cost patterns for these children.

b. Target populations and patient identification, recruitment, and enrollment

The program broadly targets children enrolled in Medicaid, especially Medicaid managed care. Table II.A.1 describes specific target populations for the four patient-focused components, each of which adapted either the target population or recruiting and enrollment processes over time. For example, at NCH, program staff shifted from targeting everyone on a given unit to focusing on families with higher needs (such as families of children with first-time admissions or with a history of barriers to outpatient care). NCH made this shift to accommodate the increased demand on staff time caused by the expanded number of participating units. At ACH, the parent peer partners initially targeted all families in the inpatient behavioral health care unit, but in spring 2014 they began prioritizing work with Medicaid families due to demand and staff availability. Likewise, the NCH behavioral health care coordination program began by identifying patients from lists sent by Medicaid MCOs that identified children with behavioral health care admissions, but later in the program many referrals came from clinical and community service providers, including hospital units, community mental health providers, and the county developmental disability organization. Because of perceived success with improving care for this population, NCH and ACH expanded complex care interventions to an additional population of children with high needs—those with tracheostomies.

The target population for replication for the ACO model at ACH was approximately 212,000 children enrolled in Medicaid based on family income in 12 counties in northeast Ohio. PFK and the ACH-based ACO also planned to target an additional 25,000 children eligible for Medicaid based on disabilities, who were previously excluded from managed care. Based on challenges in negotiations with MCOs, ACH revised its ACO plan to begin with the 4 counties with the largest populations and add the 8 other counties in the future. ACH also planned to begin with the income-eligible Medicaid population and then enroll those with eligibility based on disability. Although the award funded no interventions for the existing PFK ACO population, it did fund data analytics services to help guide the PFK interventions outside the award, targeting the approximately 300,000 children in the PFK-covered region of the state.

ACH's prematurity initiative targeted obstetric practices and women at risk for repeat premature delivery. Although the program initially planned to work only with obstetrics providers, ACH attempted to improve the reach of the program by providing direct education in neonatal intensive care units to mothers who had delivered a premature infant. By spring 2015, ACH had attempted to further expand its educational efforts to include pediatricians who cared for preterm babies, who are likely to have a longer-term relationship with the mothers.

c. Service delivery protocols

Table II.A.1 (above) provides key details about the service delivery protocols for the four program components focused on patients and their families. At NCH, the frequency of contacts

with caregivers for behavioral health and complex care coordination was driven by care coordination certification requirements from the PFK contracts with MCOs. Care coordinators were required to have in-person meetings with a caregiver at least once every three months. These meetings frequently occurred during a scheduled visit with another provider, but the coordinators did perform home visits as needed. ACH initially followed these requirements for complex care coordination in anticipation of similar contracts but, in the absence of agreements with MCOs during the award, made adaptations for efficiency, such as less frequent in-person contacts. The ACH complex care team developed a tiering system, based partly on the work of other HCIA awardees, and children with more complex medical or social needs received the quarterly in-person visit and children with less complex conditions still received a quarterly contact, usually by telephone, and at least an annual in-person visit. The ACH prematurity prevention component began with educational outreach to obstetrics providers and mothers of premature infants, and then expanded to pediatric primary care providers in the third year of the award.

d. Intervention staff and workforce development

Table II.A.2 describes the variety of new staff members that NCH and ACH hired and trained for the program. Behavioral health staff included parent peer partners, care coordinators/case managers, and their supervisors. Parent peer partners were laypeople who were required to have experience with significant behavioral health problems in at least one of their own children. They went through extensive training with an outside consultant in approaches to family-centered care, management of work with mental health clinicians, and motivational interviewing. Behavioral health care coordinators/case managers were social workers with prior experience with the community mental health system, and many had experience in the child welfare system.

The complex care intervention used multidisciplinary teams of social workers, nurses, and dieticians (included at ACH only), most of whom had extensive experience in working with children with complex needs. Additionally, NCH developed a position it referred to as a quality outreach coordinator, filled by nonclinicians with bachelor's-level training and experience working in health care settings, such as working as front-desk staff or schedulers. The main prematurity initiative staff member had a background in social work.

2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators before and during Spring 2015 site visits as well as self-reported information included in NCH's quarterly self-monitoring and measurement plans.

Table II.A.2. Key details about intervention staff

Program component	Staff members	Staff /team responsibilities	Adaptations?
Behavioral health parent peer partners and care coordination	Parent peer partners	Worked directly with families to help them communicate their children's needs to providers, express their goals for treatment, and navigate the in-hospital behavioral health evaluation and treatment processes; became formal members of the in-hospital behavioral health team to communicate families' needs and concerns and encourage family-centered care from providers	Began the involvement of parent peer partners in hospital-wide advisory groups
	Parent peer partner supervisor (social workers)	Provided training and supervision to parent peer partners	No
	Social work care coordinators/ case managers	NCH – provided ongoing care coordination through telephone, email, text, and in-person contacts with caregivers ACH – provided post-discharge case management	No
Complex care	Nurse care coordinators	Provided ongoing care coordination through telephone, email, text, and in-person contacts with caregivers focused on children with most active or complex medical conditions	No
	Social work care coordinators	Provided ongoing care coordination through telephone, email, text, and in-person contacts with caregivers focused on children with most complex social issues	No
	Dietician (ACH only)	Provided comprehensive clinical nutritional assessment at program intake and ongoing nutritional management	No
	Quality outreach coordinator (NCH only)	Assisted the care coordinators in carrying out tasks such as contacting durable medical equipment companies or working directly with families to coordinate scheduling of specialist visits	Position created in second year of the award
ACO replication	ACO administrator (senior health care management professional)	ACH – led development of ACO administrative infrastructure and contract negotiation with Medicaid MCOs NCH – provided advice to ACH administrator on ACO development; contributed to design, implementation, and evaluation of program components at NCH	No
Prematurity prevention	Coordinator (social worker)	Engaged obstetrics providers and women of infants in neonatal intensive care units through educational materials and meetings; coordinated efforts with other regional and state programs	No

Table II.A.2 *(continued)*

Program component	Staff members	Staff /team responsibilities	Adaptations?
Cross-component administration and support	Program managers	Managed day-to-day operations of the program	No
	Care coordination supervisor	Managed planning and day-to-day operations of behavioral health and complex care coordination programs	Position created in first year of the award
	Data analysts	Managed collection and analysis of internal and, when available, Medicaid administrative and claims data to evaluate program implementation and impacts	No

Source: Interviews from second site visit, March 2015; document review, March 2015.

ACH = Akron Children's Hospital; ACO = accountable care organization; MCO = managed care organization; NCH = Nationwide Children's Hospital.

a. Program enrollment

The effectiveness of program enrollment varied by program component. As of December 2014, overall cumulative program enrollment was 4,596 children, 87 percent of the final program goal. Although the program did not set specific enrollment targets for the parent peer partner or behavioral health care coordination/case management interventions, program administrators believed the interventions reached their target populations. NCH had less success enrolling families of children in outpatient treatment into an initial pilot for an online therapy program that was discontinued, but reported having more success later with another online platform targeted at families of children being discharged from an inpatient behavioral health care admission.

The complex care coordination component enrolled fewer children at NCH than anticipated, but ACH enrolled more than planned. As of March 2015, NCH complex care had identified 524 eligible children compared with an initial estimate of about 600 eligible children enrolled in Medicaid, and enrolled 312 children (about 60 percent of eligible children). The ACH complex care program had enrolled about 160 eligible children compared with an initial estimate of about 100 children.

As of spring 2015, the new ACO led by ACH had not enrolled any children because of delays in negotiating contracts with Medicaid MCOs. ACH was able to establish the corporate administrative structures needed for the ACO, but the ACH team struggled to contract with the Medicaid MCOs, which they perceived as reluctant to enter into full financial risk models similar to PFK. PFK enrolled fewer children enrolled in Medicaid based on disabilities than planned because of delays at the Medicaid agency in enrolling these children into managed care.

Due to the challenges of engaging providers, ACH focused on implementing the prematurity intervention in only a few clinics. The effort was first implemented in two hospital-based high-risk obstetric clinics, but broad educational efforts continued for many more practices in the community.

b. Service measures

NCH and ACH had mixed success in meeting their broad program process and service delivery goals. In the behavioral health interventions, they generally were successful meeting goals for family contacts in hospital care settings, but were less successful in meeting their goals for follow-up and outpatient contacts. NCH consistently met its goal of having parent peer partners discuss community resources and safety planning with families within 37 days of discharge for at least 85 percent of patients. NCH met or exceeded its goal for average parent peer partner contacts per full-time equivalent (FTE) (greater than 95) for most of the second half of 2014, but saw a decrease as the parent peer partners began to be spread across more care settings, including a new inpatient behavioral health care unit at NCH. ACH did not track these measures. NCH and ACH both had modest success in getting families to complete a behavioral health impairment assessment by 30 days after discharge, ranging from 40 to 60 percent. NCH was not as successful in engaging families in outpatient therapy with an online therapy program to supplement their care, averaging only about 40 percent of its monthly enrollment of goal of 25 patients and only about 35 percent of families activating their accounts within 30 days of enrollment, compared with a goal of at least 70 percent.

In the complex care component, NCH provided care coordination to 60 percent of eligible children compared with a goal of 85 percent. ACH did not report the proportion of eligible patients who received care coordination, but by early 2015 it was meeting its goals to complete a care plan for more than 90 percent of all enrolled children, complete medication reconciliation for more than 75 percent of all enrolled children, and collect weights on more than 75 percent of enrolled children who had a health care visit in a given month.

In the ACH prematurity prevention component, program staff collected data via surveys of mothers with a newborn in several neonatal intensive care units and special care nurseries to assess progesterone delivery rates in the community. Program administrators had initially planned to use Medicaid claims or vital statistics, but were unable to obtain these data. ACH found that throughout the award period about 50 percent of mothers with an indication to receive progesterone had received it during their latest pregnancy.

c. Staffing measures

Over the duration of the award, the program reached 95 percent of its original projection for new hires (about 45 of 48 projected FTEs). Although both NCH and ACH experienced staff turnover, both were largely able to refill positions. Staffing decreased toward the end of the award as some positions moved to in-kind support from the health systems. As the program looked to expand some interventions toward the end of the funding period, it had several open positions for parent peer partners and care coordinators.

d. Program time line

NCH and ACH implemented most of the planned interventions according to their original time line, with the notable exceptions of the ACH ACO and behavioral health telemedicine. ACH faced delays in implementing the ACO due to ongoing negotiations with Medicaid MCOs. Part of this delay resulted from the state's July 2013 Medicaid managed care reprocurement,

which changed the state's managed care model from multiple regional MCOs to five statewide MCOs. The primary barrier was reluctance from Medicaid MCOs to negotiate risk-sharing agreements similar to PFK's. Possible explanations based on our interviews with ACH staff include (1) reluctance to negotiate regional risk sharing with ACH when there were several competing hospitals serving children in the region, (2) absence of Medicaid requirements at the time for MCOs to enter into alternative payment models, (3) concerns from MCOs about becoming just "pass through" organizations without direct roles in care management, (4) lack of strong financial incentive given relatively low costs of children compared with adults, and (5) limited MCO capacity given recent expansion from regional to statewide MCOs. Additionally, the planned enrollment of children with disabilities into Medicaid MCOs did not begin until July 2013 and was slow to progress, reducing NCH's and ACH's ability to expand the ACO model to include this population. The behavioral health telemedicine intervention never moved past a few pilot sites due to delays in deploying telehealth technology related to unclear internal and external guidance about the use of equipment purchased with award funds for billable services.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.3 summarizes the major facilitators of and barriers to NCH's implementation effectiveness in each domain.

Table II.A.3. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Adaptability of program components • Frontline flexibility in implementing the program(user control)	• Adaptability
Implementation process	• Self-monitoring and quality improvement • Resources available to the program	• Obtaining population-level data for self-monitoring • Stakeholder engagement
Internal factors	• Team characteristics • Leadership characteristics • Organizational culture	• Organizational culture
External factors	• Connections to a broad professional network • Medicaid and general health care policy environment	• Medicaid and general health care policy environment • Patient needs and resources

Sources: Interviews from second site visit, March 2015; document review, March 2015.

a. Program characteristics

Two program characteristics of the NCH program had the most substantial contributions to the ongoing implementation: (1) the adaptability of the program components and (2) frontline users' flexibility in implementing the program (user control). First, NCH and ACH were able to adapt various components of the program to different settings and challenges. For example, the

parent peer partner program was based on an outpatient peer support program at Columbia University for adults with serious mental illness, which NCH and ACH adapted for children in their care settings. NCH began the parent peer partner intervention in an outpatient behavioral health crisis unit and has since adapted it to referrals from inpatient medical units caring for children with behavioral health problems, outpatient behavioral health providers, and an inpatient behavioral health care unit. Similarly, ACH initially implemented its parent peer partner program in a hospital emergency department-based crisis intervention unit and then adapted the model for its own inpatient behavioral health care unit to more directly help achieve its goal of reducing behavioral health-related readmissions. For the parent peer partner intervention, program administrators had to adapt program staffing and management policies. Parent peer partners were specifically recruited and employed on the basis of their experiences managing the mental health problems of their own children, and thus were very likely to have personal crises arising from their children's needs. The program had to adapt standard hospital staffing models and management policies to provide flexibility when these crises arose. As another example, NCH noted the low uptake of the behavioral health online therapy program among families in outpatient programs. The program then shifted focus to in-person enrollment of children being discharged from behavioral health admissions and found more family engagement with a similar online therapy tool and enrollment in the behavioral health care coordination component.

Second, program administrators provided component leaders and frontline staff significant user control, allowing freedom to refine program processes and to identify and fix problems as they arose. For example, parent peer partners described changing the timing and location of administering a follow-up behavioral health impairment scale to be less intrusive for families. In addition, the nurses, social workers, and dieticians in the ACH complex care coordination teams described multiple rounds of revisions to the process through which they assigned patients to tiers that were the basis for the frequency and mode of contact between staff and families.

Although the program successfully adapted some of the program components, it has faced significant barriers in adapting other interventions. NCH and PFK program administrators noted concerns about how the administrative structure of the HCIA limited their ability to adapt new interventions to meet their overall program goals. For example, they identified new opportunities for cost reductions and care improvements for the PFK population, such as through the management of psychotropic medications, but they did not move forward with interventions within the award due to challenges in getting approval from CMMI for new interventions through the award.

b. Implementation process

Two implementation process factors facilitated the implementation of the program: (1) self-monitoring and quality improvement activities and (2) the resources available to the program. First, NCH and ACH developed a data-oriented implementation approach. Despite reporting initial struggles with determining metrics, they developed measures for each of the program components and used them to track progress, assess success, change what was not working well, and make decisions about which program aspects to sustain. NCH took advantage of the experience that PFK and its own large research institute already had in collecting and analyzing

program data, and helped ACH leverage these resources and experience. Program leaders emphasized the importance of using a combination of program, hospital billing, and Medicaid data to monitor and adjust the program, demonstrate patients' outcomes and financial results to departmental and hospital administrators, and support the inclusion of intervention services in hospital budget and operational plans.

Second, stakeholders interviewed described several ways in which program resources supported implementation and how they were able to adapt the program to match the available resources. For example, frontline program staff reported that hospital staff not affiliated with the program were willing to share their knowledge and expertise, which was an important resource when learning how to create their roles and do their jobs. Hospital administrators were willing to provide in-kind resources to support the program. ACH was able to work with an internal expert in organizational behavior to help program administrators and staff manage behavioral change for themselves and stakeholders for their different components and target populations.

NCH and ACH also faced at least two key implementation process factors that were barriers to program implementation, including (1) obtaining population-level data for self-monitoring and (2) stakeholder engagement. First, although both NCH and ACH have been able to obtain and use internal process and billing data to monitor performance, they faced major challenges in obtaining usable state Medicaid and vital statistics data to evaluate program effects on outcomes and make comparisons to other regions of the state. These challenges included administrative hurdles in obtaining the data as well as numerous problems with data quality. As a result, NCH and ACH had to modify their data collection activities and redefine their performance metrics. For example, due to problems with Medicaid data, they have been unable to use claims to measure rates of outpatient follow-up after behavioral health-related hospital discharges. To overcome this problem, in early 2013 they began surveying families of children with behavioral health-related admissions after discharge.

Second, ACH faced challenges with engaging several key stakeholder groups. Throughout the first two years of the award, Medicaid MCOs were reluctant to implement fully or partially capitated contracts with ACH to support the development of their ACO. By the final year of the award, ACH revised their plans and began to negotiate with Medicaid MCOs for a financial model based on shared savings plus quality bonuses. ACH also faced challenges engaging obstetrics providers in the prematurity prevention component, largely because smaller practices care for relatively few of the high-risk mothers the program was targeting and the larger provider groups were involved in ongoing merger activities between hospitals, which limited their willingness to participate. Also, many larger practices began participating in a similar statewide initiative to reduce preterm births.

c. Internal factors

Characteristics of the organization implementing a program can influence implementation effectiveness. Two internal factors had notable influences on implementation of the NCH program: (1) team characteristics and (2) leadership characteristics. First, characteristics of the frontline program staff teams have been beneficial to implementation. For example, in the complex care component, an interdisciplinary approach helped the teams develop the program

and address families' needs from multiple perspectives, and the teams have worked well together to problem-solve issues. A staff member at NCH noted that being embedded in the palliative care team provided a large peer group with expertise in complex, chronic, and life-threatening conditions upon whom they could rely for ideas and additional support to care for families.

Second, hospital and health system leaders at NCH and ACH actively supported the program and discussed the need to transition to care and financial models that support population health and value-based care. Partly in response to the award, ACH began to shift its internal compensation models for providers to include quality components in addition to volume. This perspective has also led NCH to make sizable new investments into behavioral health despite a history of direct financial losses on these services; the organization's leadership views them as services that have potential to improve health and decrease costs in other aspects of care.

One internal factor, organizational culture, acted as both a barrier to and facilitator of implementation. NCH and ACH had early challenges with overcoming existing organizational culture to promote redesign and use of innovative models of care. For example, program staff and administrators reported that some behavioral health clinicians at NCH and ACH were initially very skeptical about working with new frontline staff, including parent peer partners and care coordinators. Respondents reported that as clinicians have worked more frequently with these new staff, they have become more accepting and in some cases have begun actively to seek their opinions as care team members, but this challenge has continued as the hospital hires new clinicians who are not as familiar with these interventions. In the complex care initiatives, program staff had to invest significant time in bringing together a large number of specialist clinicians within specialties, who were traditionally not interested in aligning approaches to a clinical procedure or condition, to agree to standardized approaches for care of children with feeding tubes, but felt that their persistence had resulted in significant progress. However, for both initiatives, NCH and ACH described a shift in cultural mindset over the course of the award toward more population health and family-centered care and acceptance of new staff and approaches to care to meet these goals. Administrators and program frontline staff alike noted that this cultural change required patience and persistence, but training and preparation of existing staff regarding the roles of the new staff could have helped with quicker integration of the new staff.

d. External factors

Two external factors were key facilitators of implementation: (1) connections to a broad professional network and (2) the Medicaid and general health care policy environment in Ohio. First, staff from both the behavioral health and complex care components highlighted the importance of having a broad network of contacts for helping families, including primary care providers, community and governmental agencies, managed care plans, and durable medical equipment companies. Parent peer partners and behavioral health care coordinators described assisting families access a broad set of resources to promote their ability to care for their children, including working directly with schools and connecting parents to mental health resources for themselves. NCH described working with other pediatric health systems around the state to advocate for better access to Medicaid data to use in population health management.

Second, several Ohio Medicaid policies facilitated the development and implementation of the award. Before the award, the state passed legislation promoting pediatric ACOs and, in 2014, Ohio passed legislation authorizing Medicaid to cover telehealth services. The state Medicaid agency also added a payment adjustment for care coordination to the capitation rate for Medicaid MCOs. As a result, PFK proposed a subcapitation rate to its ACO contract to support care coordination activities. This relocation of care coordination activities from MCOs to PFK was one of the goals originally established for the PFK expansion. In addition, the state began to move toward episodic payment models but planned to provide waivers for other value-based models, such as the PFK ACO model. Medicaid MCOs changed their policies to pay for progesterone, which helped support the goals of the prematurity prevention intervention by decreasing barriers to providers pursuing progesterone for their patients and patients receiving it.

At the same time, the Medicaid and general health care policy environment in Ohio and patient needs and resources were barriers to implementation. Several developments in Ohio's health care policy environment have posed significant barriers to implementation of some program components. The state underwent a reprocurement process with Medicaid MCOs in 2013, which delayed when ACH could begin ACO contract negotiations with managed care plans. Based on our discussions with awardee staff, MCOs were reluctant to enter into risk-sharing agreements with ACH for a variety of potential reasons. In addition, the state began a process of moving children who qualify for Medicaid based on a disability into managed care plans. This change created significant opportunities for care improvement and cost reduction in the PFK population, but the transition has occurred much more slowly than the state originally planned. Also, original estimates of the population of children eligible for Medicaid based on disability to be enrolled into managed care were overstated because the state elected to exclude children in waiver programs; this was problematic because these children constitute a significant portion of the complex care intervention's target population. Thus, relatively few patients targeted by the complex care intervention were enrolled in PFK through managed care. This limited the potential for cost savings in the PFK-enrolled population during the award period.

Second, the extent of patient and family needs in a Medicaid-enrolled population presented ongoing challenges for the program. In the behavioral health and complex care coordination interventions, care coordinators described having to try to address a wide variety of socioeconomic, educational, and parental needs in order to improve the health care and health of the children. For example, some care coordinators described having to attend individualized educational planning meetings for children to try to optimize supports and to connect parents with supports for their own mental and physical health care needs so they would be able to better care for their children. Behavioral health care coordinators also described frequently having to enroll a sibling of a target child due to similar behavioral health needs, which could result in the care coordinators reaching their maximum case load more quickly than anticipated.

4. Sustainability and scalability

NCH and ACH have undertaken several steps to sustain, and in some cases scale up, their interventions beyond the period of the HCIA. Broadly, NCH is sustaining most of its interventions by incorporating them into the PFK ACO or having a clinical department take on responsibility. ACH has yet to establish contractual agreements to support its ACO, so short-

term sustainability plans involve scaling back most interventions and obtaining support through clinical departments or short-term funding from the health system. For ACH, longer-term sustainability will largely depend on establishing new payment agreements with Medicaid MCOs. Both NCH and ACH have used service delivery and financial metrics collected during the award to demonstrate sustainability to health system and clinical department administrators. Hospital administrators expressed an expectation that measures continue to be collected to evaluate ongoing support for interventions.

In the behavioral health components, both institutions transitioned responsibility for the parent peer partner intervention to their respective clinical behavioral health departments during the award period and received commitments for ongoing institutional funding. The NCH hospital board approved millions of dollars over the next five years to behavioral health programs, making sustainability more likely. As a result, NCH anticipated increasing the size of its parent peer partner intervention in the second half of 2015 to cover current services and its newly opened inpatient behavioral health care unit. ACH planned to scale back its intervention with fewer staff. For behavioral health care coordination, NCH planned to incorporate the behavioral health care coordinators into PFK after the award and fund them through its managed care contracts. ACH planned to include funding for a behavioral health care coordinator/manager in its behavioral health institutional budget proposal. For its pilot programs, PFK planned to support the intervention linking families to an online therapy program after a behavioral health care hospitalization, but the online program targeted at families of children in outpatient programs will be discontinued.

NCH and ACH also planned to sustain their complex care coordination interventions. As with behavioral health at NCH, complex care coordination will be incorporated into the PFK ACO and supported through its managed care contracts. In anticipation of more children with complex needs enrolling into Medicaid managed care, PFK already planned to increase the number of care coordinators. NCH and PFK expect to replicate the model for other high-needs populations, such as children in foster care or with complications from prematurity. NCH's work on standardizing care protocols and patient education was also integrated into the organization's quality improvement team with plans to adapt the tools developed for other populations. ACH expected to sustain a scaled-back version of its complex care coordination activities. ACH's intervention was built from an existing palliative care program and was being integrated into that program by the end of the award. However, financial sustainability of the full intervention approach likely depends on ACH's ability to finalize risk contracts with Medicaid MCOs to fund the staff. One complex care administrator noted during the spring 2015 site visit that ACH has "given families really great case management but [after the award] it will be the Toyota version instead of the Cadillac of case management." For example, ACH will no longer help families with nonmedical contacts such as meetings with school systems. Regardless of whether ACH continues the complex care program, it planned to expand lessons learned to the larger palliative care program, such as implementing the tiered system for patient follow-up.

In the prematurity prevention component at ACH, the main staff member was approved for a position funded by the hospital, but longer-term sustainability will likely depend on obtaining

contracts with Medicaid MCOs that would make cost-savings from preventing premature births beneficial to the hospital.

Outside the award, another large children's health system in Ohio (Cincinnati Children's Hospital Medical Center) began implementing an ACO model with risk-sharing arrangements modeled on PFK with two of the five Medicaid MCOs in the state. NCH and PFK were negotiating with the state for care coordination fees for children not enrolled in Medicaid managed care because many children with disabilities had been maintained in waiver programs. NCH and PFK also were developing plans to develop ACO models with commercial payers.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

In its original application for HCIA funding, NCH proposed to improve health and health care and lower costs for children enrolled in Medicaid by enhancing an existing pediatric Medicaid ACO model and spreading the model to new geographies and populations. To achieve these goals, NCH sought to implement new interventions to support the existing PFK ACO, implement the same interventions at ACH, and support the development of a new ACO led by ACH. After nearly three years of HCIA funding, NCH and ACH have largely succeeded in implementing new interventions to improve the care of children with complex medical and behavioral health care needs, but the program was less successful in developing a new ACO led by ACH. Program implementation was facilitated by adaptability to the needs of families and program staff, a program culture of self-monitoring and quality improvement, and the development of interdisciplinary teams. In contrast, implementation was hindered by challenges in obtaining usable population level Medicaid claims data, engaging a broad range of specialty providers in efforts to standardize care processes, and limited willingness by Medicaid MCOs to develop capitated payment arrangements with ACH. Despite this, NCH, PFK, and ACH plan to sustain all the core interventions for at least the near future.

Our next steps for this evaluation are to (1) monitor ongoing program implementation and plans for sustaining the program beyond the no-cost extension funding period by reviewing quarterly data submitted by UHC; (2) evaluate trainee attitudes and experiences with the program in the third year of the award through an administered survey; (3) complete agreements to obtain Ohio Medicaid data and perform an impact evaluation during the final option year of the evaluation contract; and (4) synthesize implementation, survey, and impact evaluation findings to assess the success of the NCH program in meeting its goals to improve health care for children enrolled in Medicaid in Ohio.

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

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Rutgers Center for State Health Policy

March 2016

Katharine Bradley
Purvi Sevak
Cara Stepanczuk
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

RUTGERS CENTER FOR STATE HEALTH POLICY

This individual program report provides a summary of the findings to date from our evaluation of the primary care redesign (PCR) program implemented by the Center for State Health Policy (CSHP) at Rutgers University under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the CSHP program. Section II summarizes the evaluation findings. We first assess the effectiveness of program implementation (Section II.A) and then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on participants' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF CSHP

CSHP received a three-year, $14.3 million award to implement a community-based care management program at four provider organizations. Table I.1 summarizes key features of the program. Based on the Camden Coalition of Healthcare Providers (Camden Coalition) care coordination model, the CSHP program used multidisciplinary, community-based care teams to connect participants who are frequent users of hospital services ("high utilizers") to appropriate clinical and social services, help them manage their conditions, and overcome socioeconomic obstacles to care. The CSHP program aimed to decrease unnecessary hospital admissions and participants' use of emergency department (ED) visits, improve health outcomes, and reduce the average annual cost of care by 14.8 percent. CSHP's HCIA award ended in June 2015.

Table I.1. Summary of Center for State Health Policy at Rutgers University PCR program

Program feature	CSHP program
Award amount	$14,347,808
Implementation date	January 2, 2013
Award end date	June 30, 2015
Program description	Implementation of a community-based care management program at four provider organizations, including: • Providing care management services through multidisciplinary, community-based care teams • Graduating participants from program and transitioning them into medical homes
Innovation components	Care coordination/care management
Intervention focus	Individual
Workforce development	Hired nurses, social workers, community health workers, behavioral health providers, and other clinical and nonclinical staff, depending on the implementation site, as well as program managers to supervise teams and improve operations
Target population	High utilizers of inpatient services; patients with chronic conditions
Program setting	Community-based
Market area	Local (4 states)
Market locations	Allentown, Pennsylvania; Aurora, Colorado; Kansas City, Missouri; San Diego, California
Core outcomes	• 14.8 percent reduction in total annual cost of care through decreased hospital admissions and ED visits

Source: Review of CSHP program reports, March 2015.

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how it was adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss findings related to program sustainability and scalability.

We based our evaluation of CSHP's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics (aggregated across all four CSHP sites), telephone discussions and follow-up communications with program administrators (for aggregate and site-specific information), and information collected during our second round of site visits conducted in April 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports. We visited two CSHP sites in 2015, one based at a Federally Qualified Health Center (FQHC) in Aurora, Colorado, called Metro Community Provider Network, and one based at an independent physician association (IPA) in San Diego, California, called MultiCultural Medical Group. We visited CSHP's other two sites (in Allentown, Pennsylvania, and Kansas City, Missouri) during the first round of site visits in 2014; we included information gathered at those sites in the summary of the CSHP program in Mathematica's first annual report (Gilman et al. 2014).

1. Program design and adaptation

a. Program component

The CSHP program had one component: care management and care coordination of participants with multiple chronic conditions and high utilization through multidisciplinary, community-based care teams. Each of the four sites implemented the program differently to conform to the organizational, cultural, and financial characteristics of the sites' host institutions and to accommodate the views of important local stakeholders, such as hospitals. Table II.A.1 describes how the two sites we visited in April 2015 implemented and adapted the intervention.

b. Target populations and patient identification, recruitment, and enrollment

Table II.A.1 provides key details about the target populations and participant identification, recruitment, and enrollment processes. As the table shows, CSHP directed its intervention services mainly to participants who had high rates of hospital service utilization and conditions that would benefit from improved care management, although specific eligibility criteria varied across sites. Both sites visited in April 2015 adapted their eligibility criteria over the course of the program to enroll more patients.

Table II.A.1. Key details about program design and adaptation

	Community-based care coordination/care management	
	FQHC-based site	**IPA-based site**
Target population	Patients who had high rates of hospital service utilization and conditions that would benefit from improved care management; patients with three or more inpatient stays within a six-month period; expanded to include patients with multiple ED visits and uninsured patients.	Patients who had high rates of hospital service utilization and conditions that would benefit from improved care management; patients with two or more inpatient stays within a six-month period; expanded to enroll patients on a case-by-case basis, such as those with two admissions within an eight-month period or those with active substance abuse.
Patient identification	Partnered with a nearby academic medical center; the hospital's electronic health record (EHR) system generated a flag for ED patients who met criteria for admission to the program. Care team members located in the ED then interviewed patients and excluded them for behaviors or conditions that meant they might not benefit from the program. For example, the care teams excluded those who demonstrated drug-seeking or violent behaviors in the hospital, those who were intoxicated during the initial interview, and those who did not express an interest in the program.	Relied on patient referrals from providers within the IPA system, as well as referrals from health plans that contract with those providers, local hospitals, and graduated patients. Also used referral information to exclude patients whose needs were likely to exceed the program's resources and patients for whom the program was unlikely to make a difference in hospital admissions or ED visits, such as patients with sickle-cell disease.
Patient recruitment and enrollment	Care team members recruited patients who fit enrollment criteria while patients were in the ED. Enrollment was completed at the first patient visit, usually in patients' homes.	Care team members recruited patients by telephone. If patients consented to enrollment, enrollment was completed at the first patient visit, usually in patients' homes.
Service delivery protocol	Care team members traveled to participants' homes or other participant-requested locations to provide assessment and treatment, and returned to the host organizations to coordinate their efforts because the site served as participants' main source of primary care during the intervention. Care management and coordination included both medical and behavioral health services, such as brief interventions for depression or anxiety and prescriptions for psychiatric medications. Care team members used a checklist to ensure participants met goals for graduation from the program and assessed readiness for graduation monthly. Every member of a participant's care team reviewed the checklist. At graduation, advance-practice nurses on the care team handed participants off to a permanent primary care provider. The care team entered information on a participant's needs into the EHR to communicate with clinicians who took over the participant's care. This site designed its intervention as a 60-day program and largely kept to that time frame.	Care team members traveled to participants' homes or other participant-requested locations to provide assessment and treatment, and returned to the host organizations to coordinate their efforts. Nurses on the care team provided basic medical services and accompanied participants to visits with their primary care providers. Care teams were relatively small and made in-person visits only as needed. The site relied on volunteers to make weekly telephone calls to check on participants. The lead care team member for each participant made a presentation to the team describing the participant's progress toward his or her goals to assess readiness for graduation, which the team assessed approximately monthly. To support the graduation decision, staff developed a participant engagement measure (a scale from 1 to 10) and applied it at enrollment and at regularly scheduled intervals. When staff observed reductions in engagement, they asked participants to sign a participation contract that clarified the participant's responsibilities and program parameters. The care team also informed primary care providers of participants' graduation and need for additional support.

Table II.A.1 *(continued)*

	Community-based care coordination/care management	
	FQHC-based site	IPA-based site
Adaptations	Yes; after implementation, the target population expanded; behavioral health screenings became mandatory soon after program launch; and the timing of visits by different members of the care team changed to meet the needs of participants, who were overwhelmed when large teams conducted home visits.	Yes; after implementation, the site modified the target population criteria. The treatment period was intended to be 90 days but had been extended to six months or more for many participants, resulting in high caseloads. In response, the site refocused care team meetings on participants' goals and the potential for graduation. The site also created a participant engagement scale to help assess whether to transition disengaged participants out of the program without graduating them.

Source: Interviews from second site visit, April 2015; document review, March 2015.
FQHC = Federally Qualified Health Center; IPA = independent physician association.

To identify participants, the FQHC-based site relied on data from a nearby academic medical center and the IPA-based site relied on patient referrals from a variety of sources. After identifying participants, both sites excluded those who would not benefit from the program or whose needs exceeded program resources. If sites determined that potential participants would benefit from the program, they proceeded with recruitment and enrollment activities. At either site, if a recruited patient consented to participate, a member of the care team conducted an initial visit to complete enrollment. This usually occurred in the participant's home.

c. Service delivery protocols

Like the Camden Coalition model, the CSHP program employed community-based multidisciplinary care teams to help program participants stabilize their medical and social conditions. Although the CSHP mobile care teams worked with participants in the community, the teams were housed within an existing health care provider organization at each of the four implementation sites. The provider organizations hosted the teams and supported their program operations by providing meeting space, program oversight, advocacy among community partners, and other resources. However, the HCIA-funded program staff, data systems, and protocols were not fully integrated within the host provider organizations; the care teams conducted their HCIA-funded intervention services independently from their host providers' normal operations.

CSHP encouraged each implementation site to develop its own service delivery protocols using the Camden Coalition model as a guide. With the goal of customizing care plans and care coordination/care management services to participants' needs, both sites visited in 2015 implemented flexible service delivery strategies and workflows. Table II.A.1 highlights key details about the service delivery protocols for both sites.

The model pioneered by the Camden Coalition aimed to provide support to participants for 60 to 90 days in most cases. To assess participants' readiness for graduation, care teams at each

site considered participants' progress toward their goals, whether their medical and social needs were stabilized, and whether they were prepared to navigate the health system without assistance. Care teams at both the FQHC- and IPA-based sites made decisions to graduate participants by consensus. Notably, both sites instituted post-graduation services mid-way through the HCIA-funded period. The FQHC-based site created an entirely new "graduate clinic" where program graduates continued to receive care from the medical providers on the intervention care team. The site created the clinic after observing that some graduated participants had difficulty transitioning to new providers. The care team also offered coordination services to past participants of the program to help them navigate short-term obstacles. The IPA site developed post-graduation services that focused on care management rather than providing medical services. Specifically, the IPA deployed volunteers to engage graduated participants in addition to current program participants. As one member of the care team explained during our second round of site visits, "Keeping a touch on [the graduated participants] is the thing that will keep them from regressing." Volunteers made weekly calls, and sometimes in-person visits, to reinforce lessons learned during the program and manage needs that emerged over time, alerting the care team if graduated participants required extra assistance.

d. Intervention staff and workforce development

The structure of the care teams varied between the FQHC- and IPA-based sites, as shown in Table II.A.2. Both sites developed care teams to include some staff who focused on meeting participants' medical needs, and others who focused on social needs. The care team at the FQHC-based site underwent many changes over the course of the HCIA-funded period. Early in the intervention, the site was adequately staffed to operate three care teams. Staff turnover increased as the end of the HCIA funding period approached due to uncertainty about the program's future. These staffing vacancies caused the site to consolidate the members into a single team, with different staff taking turns serving participants. The site also employed a program manager and director. At the beginning of the HCIA-funded period, the site employed a co-director model that included a behavioral health services executive from the FQHC, but realized that the intervention was more efficient with a single director. The behavioral health executive continued to serve on a steering committee for the site.

At the much smaller IPA-based site, the HCIA funded only two of the four care team positions; after experiencing high caseloads, the site secured a separate foundation grant and doubled the original two-person care team. A medical director, program director, and program manager provided clinical and organizational support oversight, and a group of volunteers supported the care teams. During our second round of site visits, a program administrator described the volunteers as the program's "secret sauce."

CSHP conducted informal training and workforce development activities at both sites. At the FQHC-based site, health coaches also received formal training specific to their roles in the program, including first aid, mental health needs, health coaching, and motivational interviewing. At the IPA-based site, program administrators developed a formal training module for volunteers and new care team members received hands-on training that involved shadowing other staff.

Table II.A.2. Key details about intervention staff

	Staff responsibilities	
Staff position	**FQHC-based site (one to three teams)**	**IPA-based site (two teams)**
Case manager (RN)	--	One nurse (RN) (of two total) served as both a nurse case manager, making assignments to the teams based on caseload and patient acuity, and as a provider of medical services to participants.
Patient navigator	--	One patient navigator, a social worker by training, assisted participants with medical needs and coordinated medical services.
Community health worker (CHW)	Two CHWs, working across teams, conducted patient identification, screening and pre-enrollment activities while embedded at the site's hospital partner. CHWs reviewed inclusion and exclusion criteria in the hospital's electronic health record to identify good candidates and screen those who were not a good fit. CHWs also helped people who did not qualify for the program to connect to a PCP or other community resources. CHWs initially attended first home visits with clinical care coordinators, but managers refocused positions to concentrate on enrollment.	Two CHWs, one per team. One of these was an RN by training, providing medical services and accompanying participants to physician appointments. One was a social worker by training.
Social worker	Three social workers served as clinical care coordinators**,** attending the initial home visit to screen for benefit eligibility, and identified resources at the community, state, and federal levels. The social workers engaged participants on a weekly basis (or more frequently) to provide basic case management for participants. They also coordinated services of other team members and conducted the last home visit before graduation.	--
Health coach	Three health coaches used motivational interviewing techniques and direction from medical providers to develop personalized plans for participant-directed health improvement, such as diabetes self-care.	--
Medical assistant	Three medical assistants supported NPs during home visits and in the graduation clinic, by performing services such as drawing blood.	--
Nurse practitioner (NP)	Two NPs served as the primary medical providers for enrolled participants, with support from medical assistants and health coaches. The NPs often made the second home visit to conduct screenings, medical history, and a physical.	--

Table II.A.2 *(continued)*

Staff position	Staff responsibilities	
	FQHC-based site (one to three teams)	IPA-based site (two teams)
Behavioral health provider	Two LCSWs and one psychiatric NP provided behavioral health screenings for every participant. LCSWs provided one-on-one counseling as needed, including short-term behavioral health interventions, and referred participants to a partnering mental health clinic for longer-term needs. The psychiatric NP prescribed psychiatric medication.	--
Program manager	One program manager supervised day-to-day program operations and made suggestions to improve workflow, with support from the program director.	One program manager supervised day-to-day program operations and made suggestions to improve workflow, with support from the program and medical directors. The site created this role after the team realized the need for additional systems and processes, such as standardized participant profiles.
Data specialist	One data specialist completed mandatory program reporting and monitored outcomes at the request of the program director, with support from a data support technician employed by the FQHC.	One data specialist assisted the program director with mandatory program reporting and outcomes monitoring.
Volunteers	--	Volunteers made weekly telephone calls to enrolled participants and monthly calls to graduated participants to issue reminders and check for needs that might require the attention of the care team. The site developed the volunteer program mid-way through the HCIA-funded period to augment scarce staff resources.

Source: Interviews from second site visit, April 2015; document review, March 2015.

Notes: Many support and supervisory staff members at both sites (such as data specialists and program directors) were partially funded by the cooperative agreement and partially funded through in-kind donations, although the proportions of HCIA funding and in-kind donations for each position changed over time.

Numbers of care team staff presented in the table reflect fully staffed sites.

CHW = community health worker; LCSW = licensed clinical social worker; NP = nurse practitioner; PCP = primary care provider; RN = registered nurse.

2. Implementation effectiveness

This section examines the evidence on implementation effectiveness. We evaluate implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators and self-reported information included in CSHP's quarterly self-monitoring and measurement reports. Table II.A.3 summarizes the self-reported measure targets and achievements of CSHP's program implementation.

Table II.A.3. CSHP self-reported program implementation measures

Measure	Target	Actual	Met target?	Adaptation?
Program enrollment	1,691	1,011 (as of March 2015)	No	Yes, expanded target population for care coordination program, as noted in Table II.A.1
Participant encounters	Not specified	9.6 encounters per month (as of March 2015)	n.a.	Yes, the IPA-based site added non-HCIA-funded care team staff and volunteers and increased the frequency of telephone contacts with participants
Initial home visits within 7 days of discharge	Not specified	77 percent of participants with initial home visits within 7 days of discharge	n.a.	No

Source: Interviews from second site visit, April 2015; document review, March 2015.

Note: These measures pertain to the entire CSHP program, including all four implementation sites.

n.a. = not applicable.

a. Program enrollment

Across its four implementation sites, CSHP enrolled 1,011 people through March 2015, 60 percent of its overall target of 1,691 enrollees. Enrollment success varied by site; for example, the IPA-based site was within 10 enrollees of reaching its target of 160 at the time of our site visit (April 2015); the FQHC-based site reported a final enrollment total of 600, 67 percent of its initial enrollment target of 900. Of the 1,011 enrolled participants across all four sites, roughly half graduated from the program and 32 percent exited the intervention for other reasons (such as moving out of the catchment area, becoming unreachable by care team staff, declining to participate further, or death). The remaining 11 percent were still active participants as of March 2015. Administrators did not set goals for graduation rates.

Across all four sites, CSHP reported that a majority (71 percent) of program participants had three or more chronic diseases at enrollment. The most common chronic diseases included asthma, chronic obstructive pulmonary disease, depression,diabetes, heart failure, hyperlipidemia, and hypertension. Although the program did not target patients with specific chronic conditions, this patient profile is in line with expectations for a program that targeted high utilizers.

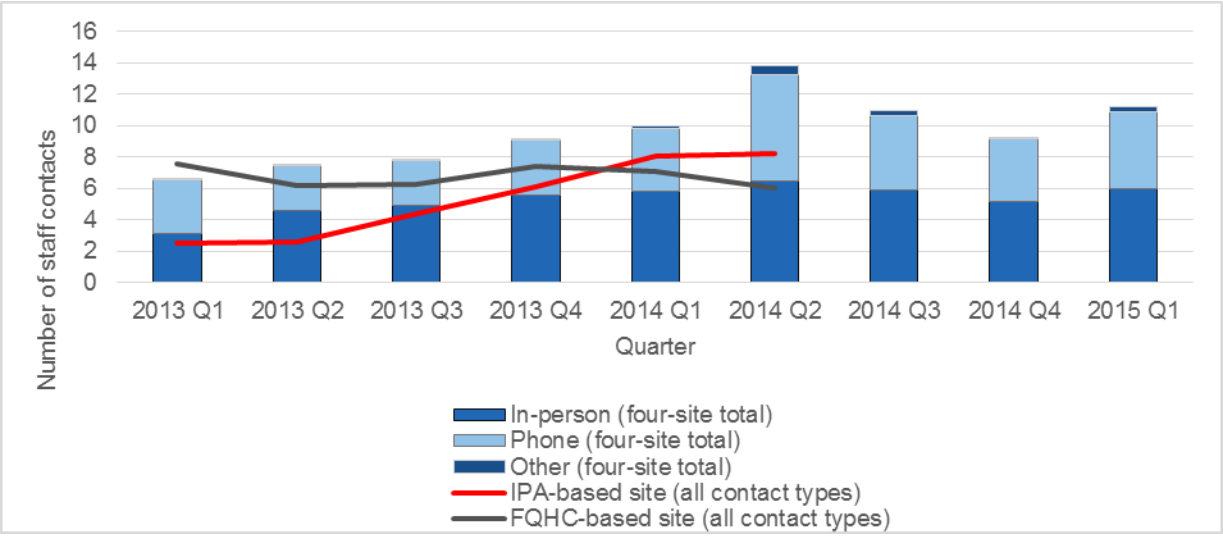
b. Service measures

CSHP's sites generally followed the Camden Coalition model of frequent, in-person contact after hospital discharge under the belief that engaging participants quickly and consistently increased the likelihood that they would achieve their self-identified goals. We use two measures to describe the degree to which the program adhered to these principles: (1) participant encounter hours and (2) timely initial home visits.

First, care teams across all four CSHP sites had an average of 9.6 encounters per participant per month for an average of six hours per month. The IPA-based program initially had fewer

contacts with participants due to a small team size (see Figure II.A.1). After adding non-HCIA-funded care team staff and volunteers, and increasing the frequency of telephone contacts with participants, the IPA site was able to increase the number of participant contacts from 2.5 per participant-month in the first quarter of operations (2013 quarter [Q] 1) to 8.2 per participant-month in (2014 Q2), which was closer to the average across all sites (9.6 per participant-month) in 2015. In contrast, the FQHC-based site's administrators felt that they were overstaffed in the early implementation phases, which enabled them to provide services to patients who were not eligible to participate in the program in addition to enrolled participants. As the program ended, staff attrition limited the site's ability to provide home-based services. On average across all four sites, care teams spent 87 percent of the time they engaged in care management on in-person visits and 12 percent of that time on telephone calls (through March 2015).The IPA-based site deviated from this average; the care team and volunteers, together, spent roughly half of their total participant contact time on the telephone as a method of extending staff resources and achieving a higher patient encounter rate.

Figure II.A.1. Staff contacts per participant-month



Source: Interviews from second site visit, April 2015; document review, March 2015.

Second, CSHP sites made it a priority to conduct an initial home visit soon after hospital discharge, although they did not specify a targeted time frame. According to data prepared by CSHP, as of September 2014, the four sites were able to complete the initial home visits within one week of discharge for 59 percent of participants, on average. The FQHC-based site was able to complete the initial home visits within one week of discharge for 77 percent of participants (the highest of the four sites) and the IPA-based site was able to complete initial home visits within one week for 41 percent of participants (the lowest of the four sites).

c. Staffing measures

At different times, both sites visited in 2015 experienced staffing issues that reduced their availability for participants' visits. The IPA-based site was understaffed in the first year of

operations. After the site secured additional (non-HCIA) funding, it was able to expand the number of frontline staff and add a program manager to improve organizational and staff capacity. The site also recruited and maintained a dedicated volunteer corps to augment care management services. The FQHC-based site was understaffed as HCIA funding came to an end, mainly due to staff attrition. The site also faced challenges in recruiting and retaining behavioral health staff. For example, the site had trouble filling a vacant role for a psychiatric nurse practitioner with prescribing authority.

d. Program time line

The CSHP program began enrolling participants in January 2013. The four sites implemented their programs at different speeds due to variation in administrative capacity and resources, although each site implemented its program within the expected timeframe.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external factors. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.4 summarizes the major facilitators of and barriers to implementation effectiveness in each domain at the two sites we visited in 2015.

Table II.A.4. Facilitators and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Staff perceptions of the relative advantage of the program compared with the standard delivery of care • Frontline staff flexibility in applying the model to meet the needs of individual participants • Adaptability of the program to meet the needs of participants and staff	• Rapid adaptation and frequency of changes to care team roles • Target population perceived as too narrow or too broad by different care team members at different times
Implementation process	• Monitoring progress to guide ongoing improvement • Engagement of staff • Engagement of stakeholders such as local primary care providers, health plans, community organizations, and local political leaders	• Staff turnover in the face of an emotionally demanding job and employment uncertainty at the end of award period • Insufficient training reported by some staff • Difficulty engaging hospitals
Internal factors	• Commitment of leadership within the host institutions • Team communication and cohesion	• Difficulty with communication among team members and with team cohesion • Fragmented or weak supervisory structure • Health IT infrastructure
External factors	• No significant facilitators noted	• Patients with complex needs and patient resource constraints • Limitations in the social service and health care systems

Source: Interviews from second site visit, April 2015; document review, March 2015.

a. Program characteristics

Three characteristics of the intervention facilitated program implementation: (1) staff perceptions of the relative advantage of the program compared with the standard delivery of care, (2) adaptation of the program to fit organizational contexts, and (3) frontline staff flexibility in implementing the program.

First, program administrators and care team members viewed the Camden Coalition model as offering an advantage for improving care, improving participants' health, and reducing costs associated with frequent hospital use, compared with the standard delivery of care and even with other care coordination models. In contrast to other health care settings and roles, care teams felt the program gave them the opportunity to build relationships with participants, help participants navigate the health care system, teach them to manage chronic conditions, and address social issues that otherwise presented recurring obstacles to health improvement or appropriate use of the health system. Staff at both sites reported that they viewed the time-limited nature of the intervention as essential for empowering participants, whereas ongoing care coordination services were more likely to promote dependence. In addition, home visits presented an advantage over telephonic or clinic-based care coordination because the care team could learn more about participants' needs by visiting them where they live. As one care team member noted during our second round of site visits, "We learned that there's a whole lot more going on at home than patients bring to the clinic."

Second, the CSHP program provided frontline staff with flexibility to tailor the delivery of intervention services to the needs of their participants. During regular care team meetings, program staff identified operational process issues and obstacles to participants' progress in the program. Team members then discussed potential solutions to remove barriers and help participants meet their goals. For example, care team members identified opportunities to increase the frequency of visits or extend the length of follow-up for participants with more complex or acute needs for care coordination and management assistance.

Third, both sites' efforts to adapt the Camden Coalition model to meet their local needs were an important factor in program success. Adaptation enabled the sites to conform to the organizational, cultural, and financial characteristics of the sites' host institutions and to accommodate the views of important local stakeholders, such as hospitals. Adaptation also enabled sites to make improvements in response to self-monitoring data and to bolster staff engagement by incorporating staff suggestions. Sites made changes to multiple program elements, including target populations, patient identification systems, services provided to current and graduated participants, care team structure, workflows, intervention length, program supports such as volunteers, and data systems. Program administrators and team members generally described these innovations as both necessary and positive.

Two characteristics of the intervention acted as barriers to implementation. First, although several stakeholders cited adaptation of the model as an important factor contributing to success, a smaller number of respondents reported that rapid adaptation presented difficulties. During our second round of site visits, staff members at one site explained that frequent changes to care team roles, made in response to staff turnover and caseload fluctuations, made the program feel

"chaotic." In addition, staff at both sites we visited viewed narrowly defined target populations as a barrier to implementation. Several informants at both sites believed they could help high utilizers who fell just outside of the enrollment criteria, and they felt broader parameters would have helped them recruit more participants. At one site, several staff also expressed the view that the target population was too narrowly defined by geography, because some frequent users obtained services in the program's service area but were not residents, and were therefore ineligible. The staff reporting these views believed the program had the capacity to help a more diverse population. In contrast, a minority of staff were frustrated that the program enrolled participants for whom care coordination was unlikely to change hospital use rates, due to the complexity of their needs.

b. Implementation process

Monitoring progress to guide ongoing improvement was a key factor that facilitated the implementation of the program. Both sites conducted qualitative and quantitative self-monitoring activities to guide program improvement. For example, one site developed new strategies for graduating or transitioning participants in a timely manner after learning the average time they spent in the program was longer than at other CSHP sites. This site focused a portion of each staff meeting on discussions about each participant's progress on goals and potential for graduation. Staff at the other site regularly evaluated enrollment and quality data to monitor the characteristics of the participant population and began analyzing hospital data to better understand the reasons that some graduated participants were readmitted to the hospital (this analysis was incomplete at the time of our visit).

Two factors served as both barriers and facilitators: (1) engagement of staff and (2) engagement of stakeholders. Informants at both sites discussed the importance of committed, engaged staff members to implementation success, and described their care teams as having these attributes. At the same time, the demands of the program could act as a barrier to staff retention: one site struggled with early staff turnover, attributed to the demands of in-home visits and participants' calls during evenings and weekends. Program administrators emphasized the need for careful hiring to prevent this problem. During our second round of site visits, one administrator said, "This is different from most work people do—I think it takes a unique person." Although both sites considered their care teams to be highly committed, staff turnover at one site increased toward the end of the HCIA-funded period, and engagement was uneven across the care team. Informants gave mixed reports on whether they had received sufficient training to understand their roles; they cited employment uncertainty as a reason several care team members left the program. Informants at the other site, in contrast, described relatively higher morale. These care team members were contractors rather than employees and might therefore have been more tolerant of uncertainty about the future of the program.

Second, program administrators and staff at both sites reported that their efforts to engage external stakeholders helped to support program implementation and sustain the progress they made. In particular, both sites had success engaging local primary care providers (PCPs), at least partly because both were housed in institutions with built-in PCP networks. Positive working relationships with PCPs helped to improve participant–clinician communication, increase participants' access to care, and strengthen collaboration to meet the complex needs of high

utilizer participants. One of the sites also focused on forging a relationship with insurers that contract with providers in the host organization. Although engagement varied across health plans, program administrators said health plans generally viewed the program as beneficial, and plans became the largest source of patient referrals for the site. Working relationships among members of the care team and nurse case managers at the health plans also helped to facilitate timely authorization of specialist visits and durable medical equipment (most participants at this site were insured). Finally, both sites successfully engaged a number of community organizations to help support participants during enrollment and after they graduated from the program.

At the same time, engaging hospital providers presented a challenge for both sites. Both sites we visited, whose host organizations were unaffiliated with hospitals, described difficulties in obtaining hospital data. Informants at one site made repeated unsuccessful attempts to encourage hospitals to refer new patients and provide utilization data for current enrollees. At the other site, a minority of informants described the relationship with the site's hospital partner as successful, although most characterized it as challenging. For example, program staff noted that the hospital decreased the amount of patient information it provided at referral, citing Health Insurance Portability and Accountability Act of 1996 concerns. Program administrators noted the hospital's continued resistance to providing financial support for the program. The site also faced challenges in obtaining data from hospitals other than its single hospital partner. Because patients use different EDs in the area, these data would have been valuable as a source of patient identification information.

c. Internal factors

Several characteristics of the program sites and their host organizations were important factors in implementation success. At both sites, support from the leaders of host organizations and program administrators facilitated implementation of the CSHP program. Leaders of host organizations advocated for the program among external stakeholders, built key relationships with partner organizations, and contributed expertise in preventive care to help guide program implementation and direct clinical care. One host organization also contributed specific expertise in care coordination, case management, and health education, all services that existed in different forms before program implementation. Leaders of host institutions and program administrators at both sites also made significant contributions to sustainability planning.

Team communication and cohesion were described as a facilitator at one site and a barrier at the other. Both sites emphasized that team communication was important for program implementation, although the two sites had different levels of success in this area. Informants at one site described the care team and its supervisors as extremely collaborative, but several informants at the other site described relatively greater difficulty with communication among team members and with team cohesion. Staff reporting these challenges attributed them to a fragmented supervisory structure. Citing licensure concerns, the program assigned different clinical supervisors for each type of care team member. The program had a single administrator but not a lead clinician. As one respondent noted during our second round of site visits, separate clinical supervisors created "… a few roadblocks because people have different views [and] different philosophies on the structure of the program," according to a respondent. This view was

not unanimous, however; medical and behavioral health providers on the care team described productive collaborative relationships.

Finally, staff at both sites described technology-related challenges to implementation. Although CSHP hoped that a web-based data entry system called TrackVia could serve as a standardized reporting, communication, and performance management tool, sites varied in the degree to which they integrated TrackVia into their workflows. Staff at one site had generally positive views of TrackVia, although they commented that it did not provide a place for visit notes. Staff at the other site described more significant technology challenges caused by the need to double- and triple-code program information: staff entered data into TrackVia, the EHR sponsored by the site's hospital partner, and/or Microsoft Excel spreadsheets created by program administrators. The lack of technology integration required care teams to maintain multiple record-keeping systems and contributed to inefficiencies in program operations. Staff from both sites stressed the importance of having a single, unified documentation system; a few care team members stated that more support from program administrators and more upfront training on TrackVia's features would have facilitated better integration with the care teams' workflow.

d. External factors

Two external factors that presented challenges to program implementation were (1) the complexity and variety of participants' needs and (2) environmental factors. First, many participants faced a variety of barriers to appropriate care, including lack of stable income, health insurance, legal residency, English language proficiency, knowledge of the health system and chronic disease management, stable housing, social support, and transportation. Many also had issues with cultural barriers, mental illness, substance abuse, and traumatic experiences that made stabilizing their chronic conditions more difficult. As a result, participants turned to EDs to meet their health care needs. Care teams helped participants address these issues through education and navigation services, but recognized that addressing participants' complex needs required a commitment and sufficient motivation by the enrollees. In addition, participants' issues often took longer to resolve than the intervention's time line typically allowed.

Second, environmental factors, such as limitations in the social service and health care systems, created challenges for program implementation and negatively affected participants' outcomes. Respondents found a general lack of affordable housing, insufficient transportation services, and poor access to specialty care to be the most significant environmental barriers to stabilizing participants' social and medical conditions. During our second round of site visits, an administrator mentioned that the demand for affordable housing and shelter beds overwhelmed the supply, with the result that "You resort to getting patients on a lot of lists." A care team member noted that although she could facilitate behavioral health assessments, short-term counseling, and referrals to specialists, wait times for appointments with psychiatrists were often two months or more.

4. Sustainability and scalability

Both the FQHC- and IPA-based sites planned to continue operations and expand their target populations after HCIA funding ended in June 2015. The FQHC site's sustainability plan

included three main components, focused on the short, intermediate, and long terms, respectively: (1) integrating the HCIA-funded care team into the institution's existing care coordination program, (2) finding bridge funding through grants, and (3) obtaining long-term funding through partnerships with hospitals and the state. The IPA-based site's sustainability plan focused on three separate components as well: (1) creating a foundation to host the program; (2) finding bridge funding through grants; and (3) obtaining long term-funding through contracts with health plans, hospitals, and public agencies. Table II.A.5 provides more details on each of these plan components.

Table II.A.5. Summary of sustainability and scalability plans

Time frame	FQHC-based site	IPA-based site
Short term	The site pursued financial commitments from the host institution (Metro Community Provider Network) and its hospital partner to jointly support a portion of existing program staff. As of April 2015, the site had secured a financial commitment from the host institution, but not its hospital partner. Funding from the Metro Community Provider Network enabled the program to continue beyond the HCIA funding period, but with a decreased number of staff.	In 2012, the site created the MultiCultural Health Foundation to host the program and access grant and public funding after the end of HCIA funding. As of July 2015, the foundation took over management and financial responsibility for the HCIA-funded care management program. The MultiCultural Health Foundation planned to develop accompanying programs over time, such as a wellness center and advocacy group, to improve participants' outcomes and increase participants' access to the health care system, but these were not in place yet at the time of our site visit.
Intermediate term	Site administrators planned to seek funding from local and national foundations to help sustain the program over the next several years, as they sought to create sustainable long-term funding. As of April 2015, the site had identified several potential grant sources but had not yet applied to them. Administrators expected that reducing the cost of the care team (by reducing staff) in combination with foundation funding would enable them to maintain operations as they worked toward long-term financing for the program.	Site administrators planned to seek funding from foundations and government sources to help sustain the program in the short term. As of April 2015, the MultiCultural Health Foundation had already received grant funding from two foundations to build capacity. Site administrators prioritized building administrative capacity and information systems in preparation for increasing the program's scale and broadening its service offerings.
Long term	FQHC administrators encouraged additional local hospitals, within and adjacent to their service area, to invest in the program because they could benefit from a reduction in the amount they write off to charity care. As of April 2015, the FQHC site had constructed overall cost-saving estimates, although it was still working to quantify potential financial results for individual hospitals. Administrators noted that it is difficult to translate savings into resources to support services that are not currently billable. Program leaders also pointed out that reducing avoidable ED visits and inpatient stays could cut into a major revenue stream for hospital partners.	Site administrators planned to achieve long-term sustainability through private contracts. Due to the state's Medicaid expansions and transition from fee-for-service payment systems to managed care, health plans have experienced rapid growth in enrollment and new requirements to develop interdisciplinary care teams. Site administrators planned to use the foundation as a contractor for their high-touch, community-based care coordination/care management model. As of April 2015, the site had already entered into contracts with Scripps Health Hospitals and the MultiCultural IPA to provide services to

Table II.A.5 *(continued)*

Time frame	FQHC-based site	IPA-based site
Long term *(continued)*	The site also planned to develop long-term funding streams for care management by working with legislators and state agencies. Administrators did not state a definitive time frame for this work. Specifically, administrators hoped the state will develop incentive payment systems for quality and efficient health delivery, including payments for organizations that reduce avoidable hospital and ED visits. FQHC administrators believed the state's existing payment structure for care coordination (through existing regional collaborative care organizations and accountable care organizations) could facilitate this, but that the payment amount would have to be enhanced to reflect the level of services required for participants. Administrators also hoped to create a future state budget line item for organizations that improve the health of such patients.	their frequent users (roughly 60 to 70 participants) after the end of HCIA funding. In addition, program administrators planned to pursue a state budget line item in the future, although they expected this to be difficult to negotiate.

Source: Interviews from second site visit, April 2015; document review, March 2015.

B. Description of clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected clinical sites provided important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff are limited to a small number of clinical locations and might not reflect the perspectives of clinicians practicing at other sites. In order to assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in the fall of 2014, the third year of the HCIA-funded program. Data from the survey provide additional insights into the implementation process and experience, as well as the contextual factors that might affect implementation effectiveness at CSHP sites. It is important to note that most of the clinicians who responded to the survey (82 percent) were not HCIA-funded and were not members of the CSHP program's care teams. Rather, they were clinicians who provided care to at least five HCIA program participants and might have interacted with HCIA care team members. We therefore interpret survey responses to reflect perspectives that were largely, though not entirely, external to the CSHP programs.

In this section, we report on clinicians' views of their daily work life and practice at the CSHP program sites. First, we focus on the contextual factors that might have affected program implementation, including the characteristics of the practice location, career satisfaction and burnout, and barriers to providing high quality and patient-centered care. We then present data on the alignment of clinicians' views and experiences with the overall goals of the HCIA-funded

innovation, as well as their awareness of the CSHP program and their perceptions of program effectiveness.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

The CSHP sample included 24 eligible clinicians. Among them, 17 returned complete, valid surveys, resulting in an overall response rate of 70 percent. Of the respondents, most were external to the CSHP programs. Most were physicians and the rest were nurse practitioners and a physician assistant. These clinicians predominantly practiced at Federally Qualified or other community health centers. Most reported that their primary source of compensation was either a fixed salary or a salary adjusted for performance.

Clinicians in the CSHP sample reported working in settings that were advanced in terms of health IT. Nationally, slightly more than half of physicians practice in settings with functional electronic health records (Furukawa et al. 2014), but most clinicians in the CSHP sample reported using health IT at their practice locations. Table II.B.1 shows that most clinicians in the CSHP sample used EHRs for various functionalities, including use of electronic tracking systems and patient registries, advanced functions that are not in widespread use nationally (DesRoches, Painter, and Jha 2014). Clinicians in the CSHP sample were less likely to offer patient-facing technologies such as electronic prescription refills and appointment requests.

Table II.B.1. Electronic capabilities for clinicians and patients

Survey item	Number of Respondents	Percentage of respondents
Physicians using EHR to		
Order tests and procedures	16	94%
Access laboratory results	15	88%
Prescribe medications	15	88%
Receive drug dosing and interaction alerts	17	100%
Enter clinical notes	17	100%
Track electronic referrals	--	--
Access patient registries	--	--

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

b. How clinicians experience their careers and workdays

Because clinicians were important external partners for CSHP's four sites, and might have been colocated with the CSHP care teams, their satisfaction with their overall career, level of burnout, and perceptions of their practice environments might have affected the success of program implementation and participants' outcomes. Clinicians in the CSHP sample were largely satisfied with their careers in medicine, as 94 percent of respondents reported being somewhat or very satisfied. Although the majority (76 percent) reported feeling occasional stress at work, only a small number reported symptoms of burnout.

Clinicians in the CSHP sample gave similar ratings to their workplace management. A majority of responding clinicians (70 to 85 percent) either somewhat or strongly agreed that their management team was supportive, that they were encouraged to offer suggestions and improvement, and that they had adequate opportunities for professional development. However, only half of clinicians in the CSHP sample felt that the amount of work they were expected to complete each week was reasonable. Most clinicians in the CSHP sample (70 percent) also reported spending a significant amount of time each week (25 to 74 percent of their time) doing work that could be done by someone with less training.

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. Almost half of responding clinicians strongly or somewhat agreed with the statement "It is possible to provide high quality care to all of my patients." Major barriers to providing optimal care reported by clinicians in the CSHP sample included insufficient time to spend with patients (88 percent), lack of timely information about care provided to patients by other physicians (88 percent), patients' inability to pay for care (88 percent), insufficient reimbursement (88 percent), and difficulties obtaining specialist referrals (71 percent). Again, most clinicians in the sample were not HCIA-funded and were not members of the CSHP program's care teams, although they practiced at the CSHP program sites. Therefore, these barriers might not reflect typical perceptions of clinicians on the CSHP care teams.

c. Clinicians' perceptions of care team functioning

More than 90 percent of clinicians in the CSHP sample reported working as part of a care team within their practice sites, although most of the respondents were not members of the CSHP intervention care teams described in Section II.A. Overall, respondents' perceptions of practice-based care teams was positive. Most clinicians agreed that members of practice-based care teams relayed information in a timely manner (77 percent), had sufficient time for patients to ask questions (77 percent), and used common terminology when communicating with one another (88 percent). However, fewer agreed that practice-based care team staff followed a standardized method of sharing information when handing off patients. These measures do not necessarily reflect the ways in which clinicians worked with the CSHP care teams; however, overall positive views of care team functioning might have facilitated interaction with CSHP care teams and overall CSHP program implementation.

d. Alignment with goals of PCR

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The inclusion of the extremely important category helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are simply important. The views of clinicians in the CSHP sample generally aligned with the goals and strategies of the CSHP program. Table II.B.2 presents results based on the proportion of respondents rating each of these goals as extremely important. To the extent that clinicians in the survey sample worked with CSHP care teams or with patients in the CSHP programs, clinicians' alignment with the goals of CSHP was likely to serve as a facilitator of successful program

implementation. In addition, overall receptivity to the values and strategies of the CSHP program on the part of clinicians in the larger practice context might mean that the program sites could leverage fairly high existing levels of clinician buy-in as they constructed care teams and implemented the programs.

Table II.B.2. Importance of primary care redesign goals

Survey item	Number of Respondents	Percentage of respondents
Clinicians rating each of the following as extremely important		
Increasing access to primary care	13	76%
Improving care coordination for patients with chronic conditions	12	71%
Improving patients' capacity to manage their own care	12	71%
Reducing ED visits	12	71%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

3. Awareness of program and perceived effects

CSHP's four community-based care management programs aimed to reduce inappropriate hospital and ED use and improve participants' ability to receive reliable and coordinated primary and specialty care. Although the program focused on removing barriers from the participants' perspectives and on improving disease self-management, program administrators believed that improving the capacity of local practices to help participants manage their care was critical to their long-term success. For this reason, clinicians' perceptions of the program could have moderated the effect of the program on participants' outcomes. For example, if clinicians were aware of the CSHP program at their practice sites and believed it could enhance the care they provide, they were likely to feel more invested in providing quality care for high-risk patients targeted by the intervention. However, fewer than two-thirds of the clinicians in the CSHP sample we surveyed were at least somewhat familiar with their local CSHP program, lower than the proportion who expressed strong agreement with key PCR goals.

a. Perceived effect of program on patient care

Finally, we asked clinicians who were at least somewhat familiar with the CSHP program whether they perceived favorable effects of the program on the care they provide to participants. Almost all of the clinicians who were familiar with the CSHP program believed the program had a positive effect on the quality and patient-centeredness of the care they provide, as well as on their ability to respond to patient needs in a timely way. Most clinicians familiar with one of the CSHP programs also believed the program had a positive effect on equity, efficiency, and safety. Although none of the clinicians perceived an actual negative impact of the program, a few believed the intervention had no effect on some of these dimensions of patient care or responded that it was simply too soon to tell.

4. Conclusions about clinicians' attitudes and behavior

Although the HCIA Primary Care Redesign Clinician Survey found that few respondents were familiar with the CSHP program, most of the responding clinicians were not HCIA-funded nor were they members of the CSHP program's care teams. Almost all clinicians who were familiar with the CSHP program believed the program had a positive effect on patient care. In general, respondents' views and characteristics point to favorable conditions for the CSHP program. Responding clinicians had access to advanced HIT, were satisfied with their careers in medicine, were comfortable working within a care team, and were supportive of primary care redesign goals.

C. Impacts on patient outcomes

1. Introduction

In this part of the report, we draw preliminary conclusions, based on available evidence, about the impacts of CSHP's HCIA program on patient outcomes in three domains: (1) quality-of-care outcomes, (2) service use, and (3) spending. We describe the methods for estimating impacts (Section II.C.2) and the characteristics of the treatment group beneficiaries at the start of the intervention (Section II.C.3). We next demonstrate that the treatment group was similar at the start of the intervention to the matched comparison group, which is essential for limiting potential bias in impact estimates (Section II.C.4). Finally, in Section II.C.5, we describe the quantitative impact estimates, their plausibility given implementation findings, and our conclusions about program impacts in each domain. Although the CSHP program serves Medicaid beneficiaries, Medicare beneficiaires enrolled in managed care plans, uninsured individuals, as well as Medicare fee-for-service (FFS) beneficiaries, due to limitations in available data we have analyzed outcomes only for the Medicare FFS population (including those who are dually eligible for Medicare and Medicaid). Results might not be generalizable to the full population that CSHP serves. Our conclusions are preliminary because the analyses do not yet include Medicaid beneficiaries, nor do they cover the full time period over which the intervention is expected to have an effect.

2. Methods

a. Overview

We estimated program impacts as the difference in outcomes between the treatment group and matched comparison group, adjusting for any differences in pre-intervention characteristics. We specified a limited number of primary tests for each domain before conducting any impact analyses. Each primary test defined the outcomes, population, time period, and direction of expected effects for which we hypothesize to see impacts if the program is effective. We drew conclusions about impacts in each domain based on the results of these primary tests and the consistency of the primary test results with the implementation findings.

b. Treatment group definition

The treatment group includes Medicare FFS beneficiaries who enrolled in the CSHP program from its start on January 1, 2013, through June 30, 2014. (One of the four CSHP

program sites—Truman Medical Center in Kansas City, Missouri—enrolled two Medicare FFS beneficiaries in November and December 2012, before the program start date elsewhere; the treatment group includes those two beneficiaries for completeness.) The treatment group includes those enrolled in Medicare FFS at the time they joined the CSHP program and continuously during the four previous (baseline) quarters,[1] including those who were also in Medicaid (dually eligible for Medicare and Medicaid). We limited the analysis sample to those continuously enrolled in FFS Medicare and observable in Medicare data during the baseline period[2] to make it easier to match treatment beneficiaries to comparison beneficiaries. Continuous enrollment ensured that we had a complete record of beneficiaries' service use in the year before program enrollment. The treatment group includes participants at all four CSHP program sites: Metro Community Provider Network in Aurora, Colorado; MultiCultural Primary Care Medical Group in San Diego, California; Neighborhood Health Centers of the Lehigh Valley in Allentown, Pennsylvania; and Truman Medical Center in Kansas City, Missouri. Sample sizes at each of the sites were insufficient to allow site-level analyses.

c. Comparison group definition

We used three steps to construct a matched comparison group of Medicare beneficiaries who were similar to the treatment group beneficiaries. This section describes how we constructed the matched comparison group whereas Section II.C.4 shows the balance we achieved between the two groups on the matching variables.

First, we identified a pool of *potential* comparison members among Medicare FFS beneficiaries who met the minimum claims-based criteria for CSHP program eligibility that we confirmed the treatment group met. These criteria are that they had (1) at least one of 25 chronic conditions, and (2) an outpatient ED visit or hospital discharge at some point from November 1, 2012, to June 30, 2014. We further limited this pool to those whose zip code in the Medicare Enrollment Database (EDB) indicated residence in geographic areas that either included the treatment group or were similar in size and composition to geographic areas in which the treatment group resided:

- Allentown, Bath, Bethlehem, Emmaus, Lancaster, Macungie, Nazareth, Northampton, Reading, Red Hill, and Scranton, Pennsylvania
- Adams, Arapahoe, Bent, Boulder, Denver, Douglas, El Paso, Elbert, Fremont, Gilpin, Jefferson, Larimer, Lincoln, Logan, Morgan, Otero, Pueblo, Teller, and Weld counties in Colorado
- San Diego and Los Angeles, California
- Kansas City and St. Louis, Missouri

[1] Quarters are defined relative to the enrollment date, and, because enrollment dates can vary by beneficiary, the calendar time associated with a quarter can differ across beneficiaries.

[2] Beneficiaries are observable if they are alive, enrolled in Medicare FFS (Part A and B), and have Medicare as their primary payer (including beneficiaries dually eligible for Medicaid).

Second, for each potential comparison beneficiary, we created a *pseudo-enrollment date* to approximate the date the beneficiary would have enrolled in the intervention if he or she had been in the treatment group. The pseudo-enrollment date was drawn to correspond with CSHP enrollment dates. Specifically, for each potential comparison beneficiary, we randomly added a number of days to the ED visit or discharge date to get the pseudo-enrollment date. We drew the number of days from a frequency distribution of days between treatment beneficiaries' last ED visit or discharge before program enrollment and program enrollment. If a potential comparison beneficiary was discharged multiple times from November 1, 2012, to June 30, 2014, we randomly selected one event to choose a pseudo-enrollment date. We then limited the comparison pool to those continuously enrolled in FFS Medicare and observable in claims data during the four (baseline) quarters before their pseudo-enrollment date, consistent with the treatment group.

Third, we used propensity score matching and exact matching techniques to limit the potential comparison pool to a set of matched comparison beneficiaries similar to treatment beneficiaries on observed baseline characteristics. We used the Medicare Enrollment Database and claims in the 12 to 36 months before program enrollment (treatment group) or pseudo-enrollment (potential comparison group) to develop baseline characteristics. Matching aims to reduce selection bias in observational studies by selecting comparison beneficiaries from the pool who are roughly equivalent to the treatment group across key, observable baseline characteristics. The goal of matching is to achieve baseline equivalence between the treatment and matched comparison groups on the variables in the matching process (Stuart 2010). For CSHP, we used exact matching to stratify the sample by whether the original reason for Medicare entitlement was old age or something else (that is, disability and end-stage renal disease [ESRD]). We did this because 85 percent of the treatment beneficiaries were originally entitled to Medicare due to disability or ESRD. Separately for both of these groups (old age and not), we used propensity score matching to match treatment to comparison beneficiaries on demographic characteristics, state of residence, zip code-level poverty rate, Medicare-Medicaid dual enrollment status, enrollment or pseudo-enrollment date, health status and chronic conditions, service use (ED visits, inpatient admissions, and unplanned readmissions), and Medicare spending during the 6 and 12 months before enrollment.

Within the family of propensity score matching methods, we implemented a technique called full matching to form matched sets that contained one treatment and one or more comparison beneficiaries. The important benefit of full matching is that it achieves maximum bias reduction on observed matching characteristics and, subject to this constraint, maximizes the size of the comparison sample (Rosenbaum 1991; Hansen 2004). Each treatment beneficiary was matched to up to 10 beneficiaries from the potential comparison group to creates a more stable comparison group against which the treatment group's experiences can be compared.

d. Construction of outcomes and covariates

We used Medicare claims from November 1, 2008, to December 31, 2014, for beneficiaries in the treatment and comparison groups to develop two types of variables: (1) **outcomes**, defined for each person in each quarter that they are a member of the treatment or comparison group; and (2) **covariates**, which describe a beneficiary's demographic, Medicare enrollment-related, and

health-related characteristics during four baseline quarters and were used in the regression models for estimating impacts to adjust for beneficiaries' characteristics before the period began. As noted earlier, the quarters are defined relative to the beneficiary's enrollment or pseudo-enrollment date. Control variables were measured during the baseline period to avoid the potential bias that could occur if the intervention affected both control variables and outcomes. For example, the intervention may have resulted in greater contact with the health system and earlier diagnoses of diseases and conditions, which could have affected both health-related characteristics and outcomes. If we adjust for changes in health-related status during the intervention period, we may adjust away part of the impact of the intervention. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. We calculated six quarter-specific outcomes that we grouped into three domains:

1. Domain: Quality-of-care outcomes
   a. Inpatient admissions for ambulatory-care sensitive conditions (ACSCs, number/quarter); also called "potentially preventable admissions."
   b. Number of inpatient admissions followed by an unplanned readmission within 30 days (number/quarter)
2. Domain: Service use
   c. All-cause inpatient admissions (number/quarter)
   d. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
   e. Total Medicare Part A and B spending ($/month)
   f. Medicare inpatient spending ($/month)

Four of these outcomes—all but ACSCs and Medicare inpatient spending—are outcomes that CMMI has specified as "core" for the evaluations of all HCIA programs. Our definition of the readmission measure, however, is different from CMMI's standard definition. CMMI typically defines readmissions as the proportion of inpatient admissions that end in an unplanned readmission. Instead, we analyze impacts on the *number* of these unplanned readmissions per thousand beneficiaries per quarter because this allows us to look at the total impact on readmissions across the treatment group, rather than readmissions contingent on an inpatient admission (because the intervention might also affect the number of and type of admissions).

Covariates. The covariates include (1) measures of chronic conditions based on claims in the 12 to 36 months (depending on the condition) before the beneficiary's enrollment or pseudo-enrollment date, including the number of major chronic conditions (out of 25 mostly physical health conditions) and six specific chronic conditions (Alzheimer's disease, cancer, congestive heart failure, chronic kidney disease, chronic obstructive pulmonary disease [COPD], and

diabetes)[3], (2) the number of mental health conditions (out of six); (3) Hierarchical Condition Category (HCC) scores, where the HCC score is a measure of the beneficiary's predicted Medicare spending in the following year relative to the national average (with 1.0 indicating that the predicted spending is at the national average and 2.0 indicating that it is twice that average); (4) ED visits, inpatient admissions, unplanned readmissions, and spending in the 6 and 12 months before the beneficiary's enrollment or pseudo-enrollment date; (5) number of baseline months of dual Medicare and Medicaid enrollment; (6) demographics (age, gender, race/ethnicity); and (7) the 2012 zip code level poverty rate in the beneficiary's home zip code.

e. Regression model

We used a regression model to implement a *contemporaneous differences* design for estimating impacts. For each quarter-specific outcome, the model estimated the relationship between the outcome and predictor variables. As noted earlier, quarters are defined relative to the enrollment date. Because the sample includes beneficiaries who enrolled in the program through June 30, 2014, and outcomes are measured through December 31, 2014, each treatment and comparison beneficiary is in the sample for at least two intervention quarters. The predictor variables included the covariates (described in Section II.C.2.d) and intervention quarter-specific indicator variables for whether the beneficiary is in the treatment group. The estimated relationship between the quarter-specific treatment indicator and outcomes in a given quarter measures the average difference in outcomes for beneficiaries in the treatment and comparison groups in that quarter, while controlling for any differences in outcomes associated with differences in the covariates.

We designed the model to measure differences in treatment and comparison group outcomes separately for each quarter, because it is possible that the program's impacts had changed since the beneficiary first received program services. We can also examine differences over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the estimates, allowing for statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. Appendix 2 provides details on the regression methods, including descriptions of the weights used in the model and how the regressions account for correlation in outcomes across quarters for a given individual.

f. Primary tests

Table II.C.1 shows the primary tests for CSHP, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important (expressed as a percentage change from the counterfactual—that is, the outcomes the treatment group would have had in the absence of the HCIA-funded intervention). The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust

[3] We include these six specific conditions because they were either highly prevalent among the treatment population or identified as exclusion criteria (Alzheimer's disease and cancer) for CSHP's program.

evidence about program effectiveness (see Appendix 3 for detail and for a description of how we selected each test).

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** The CSHP program aimed to reduce spending by 14.8 percent, by reducing avoidable hospitalization and ED visits. Based on this goal, we specified primary tests in three domains in which we expected the program to have an effect. In the quality-of-care domain, we expect the program to reduce ACSCs and 30-day unplanned readmissions. In the service use domain, we expect the program to reduce all-cause admissions and ED visits. Finally, in the spending domain, we expect the program to reduce inpatient hospital spending and total spending.
- **Time period.** We expected reductions in outcomes across all domains to be largest during program participation and perhaps harder to identify as the health of the treatment and comparison group members evolved. Because the length of the intervention varied to accommodate patients' needs (on average, the intervention lasted 2 to 8 months post-discharge, but it could last up to 13 months), we chose to specify our primary tests based on outcomes in the four quarters following a participant's enrollment date (that is, I1 to I4 in Table II.C.1). To implement each primary test, we take the average of the regression-adjusted estimates across the four quarters I1 to I4 for that outcome. Please note that, to specify the primary tests, we had to make an assumption about Medicaid data availability for our third annual report, due in 2016. We believe it is possible we could have Medicaid data covering the period through mid-spring of 2015 for most CSHP sites, so we specified our primary tests accordingly. However, *this assumption is intentionally optimistic*. If Medicaid data are available only for a shorter period, we will have to revisit the primary test time period for the Medicaid population.
- **Population.** CSHP's program was meant to influence outcomes across all domains for all program enrollees. CSHP identified no program subgroup as expected to have different program impacts than other enrollees. Our primary tests are specified for Medicare FFS and Medicaid beneficiaries.
- **Direction (sign) of the impact estimate.** The primary tests use one-sided tests for a reduction, relative to the counterfactual, for each of the six outcome measures across the three domains.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we have pre-specified thresholds for what we call substantive importance. The decline of 11 percent that we chose for substantive importance for total Medicare spending is 75 percent of CSHP's anticipated impact on spending. (We used 75

Table II.C.1. Specification of the primary tests for Rutgers Center for State Health Policy

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population[d]	Substantive threshold (impact as percentage of the counterfactual) [e, f]
Quality-of-care outcomes (2)	Inpatient admissions for ambulatory care-sensitive conditions (#/person/quarter)	Average over intervention quarters 1 through 4[c]	All Medicare FFS and Medicaid beneficiaries in the treatment group	-15.0%
	30-day unplanned hospital readmission rate (#/person/quarter)			-15.0%
Service use (2)	All-cause inpatient admissions (#/person/quarter)			-15.0%
	Outpatient ED visit rate (#/person/quarter)			-15.0%
Spending (2)	Medicare Part A and B and Medicaid spending ($/person/month)		All Medicare FFS and Medicaid FFS beneficiaries in the treatment group	-11.0%
	Medicare and Medicaid FFS inpatient spending ($/person/month)			-15.0%

Notes: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

[a] We adjusted the *p*-values from the primary test results for the multiple comparisons made within each domain, but not across domains.

[b] The regression models controlled for differences between the treatment and comparison groups during the baseline year when estimating program impacts.

[c] To implement the primary tests, we took the average of the regression-adjusted estimates for intervention quarters 1 through 4.

[d] To specify the primary tests, we made assumptions about the Medicaid data that will be available by our third annual report, due in 2016. We believe that we could have Medicaid data for some CSHP sites for the third annual report. However, this assumption is optimistic. If Medicaid data are available only for a shorter period, we will revise our primary test timing to cover, for the Medicare population, all months in which the HCIA program is operating (with outcomes measured through June 2015) and, for the Medicaid population, as many months as possible within this same period.

[e] For total and inpatient Medicare FFS and Medicaid FFS spending, we set the substantive threshold to 75 percent of CSHP's anticipated impacts on total spending. For the other outcomes, CSHP did not state an explicit target, and we set the threshold equal to reductions in acute-care use or spending that Peikes et al. (2011) indicated could be feasible among either high-risk beneficiaries in a patient-centered medical home program.

[f] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention.

CSHP = Rutgers Center for State Health Policy; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

percent, recognizing that CSHP could still be considered successful if it approached, but did not achieve, its fully anticipated effects). The awardee did not specify anticipated effects on the utilization or quality-of-care measures, so all of our other thresholds—for outpatient ED visits, all-cause admissions, unplanned readmissions, and ACSCs—are instead taken from the literature (Peikes et al. 2011). These thresholds are based on the assumption that a successful high-touch, community-based care management intervention could cause a reduction in spending or service use of 15 percent among a high-risk population.

Due to limitations in data availability, we were able to conduct the primary tests in this report only among Medicare FFS beneficiaries enrolled in the program by June 2014. We conducted the tests using outcomes observable through December 2014. In our third annual report, we will add outcomes for Medicare FFS beneficiaries through the end of the program (June 30, 2015), and, if data allow it, add Medicaid beneficiaries as well.

g. Synthesizing evidence to draw conclusions

Within each domain, we drew one of four conclusions about program effectiveness, based on the primary test results and the plausibility of those findings given the implementation evidence. These four possible conclusions are as follows: (1) statistically significant favorable effect (the highest level of evidence); (2) substantively important favorable effect; (3) indeterminate effect; and (4) substantively important unfavorable effect. (We could not conclude that a program had a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program unfavorable effects.)

Our decision rules for each of the four possible conclusions are described in Appendix 3. In short, we concluded that a program had a statistically significant favorable effect in a domain if (1) at least one primary test result in the domain was favorable and statistically significant, after adjusting the statistical tests to account for multiple tests (if applicable) within a domain; or (2) the average impact estimate across all primary tests in the domain was favorable and statistically significant. In both cases, we also had to determine that the primary test results were plausible given implementation evidence. We concluded that a program had a substantively important favorable effect if the average impact estimate in the domain was substantively important but not statistically significant, and if the result was plausible given implementation evidence. In contrast, if the average impact estimate was unfavorable (opposite the hypothesized direction) and larger than the substantive threshold, and unfavorable effects were plausible given the other evidence, we concluded the program had a substantively important unfavorable effect. Finally, if the tests in a domain did not meet any of these criteria, we concluded that the impact in that domain was indeterminate.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the treatment group at baseline, which can be seen in the second column of Table II.C.2. The last column of the table contains values of the variables for the national Medicare population as benchmarks.

Table II.C.2. Characteristics of treatment and comparison groups at baseline for Rutgers Center for State Health Policy

Characteristic	Treatment group (n = 115)	Unmatched comparison group (n = 217,841)	Matched comparison group (n = 1,000)	Absolute difference[a]	Standardized difference[b]	Medicare FFS average
Exact match variables[c]						
Original reason for entitlement is disability or ESRD (%)	85.2	29.1	85.2	0	0	N
Propensity matched variables[d]						
Demographic characteristics						
Age (years)	58.7	72.3	56.5	2.2	0.154	71[e]
Male (%)	46.1	40.9	45.9	0.2	0.003	44.7[f]
Race: Black (%)	53.9	11.1	52.7	1.2	0.024	10.4[f]
Zip code poverty rate (%)	27.2	14.1	26.9	0.3	0.022*	15.0[g]
Medicare-related characteristics						
Dual status (# of months in year before enrollment)	7.9	2.6	7.6	0.3	0.057	NA
Health status and chronic conditions						
HCC risk score	3.9	1.8	3.8	0.1	0.044	1.0
Chronic conditions (# out of 25)[h]	7.8	5.2	7.7	0.1	0.034	NA
Mental health conditions (# out of six)[i]	1.4	0.7	1.3	0.1	0.062	NA
Alzheimer's (%)	8.7	13.0	9.0	-0.3	-0.011	4.9[j]
Cancer (%)	6.1	12.9	5.7	0.3	0.015	NA
CHF (%)	56.5	21.7	56.2	0.2	0.006	15.3[j]
CKD (%)	62.6	27.3	61.2	0.9	0.018	16.2[j]
COPD (%)	59.1	21.7	58.5	0.6	0.012	11.8[j]
Diabetes (%)	68.7	32.3	67.2	1.5	0.032	28.0[j]
Service use and spending 6 months before enrollment						
30-day unplanned hospital readmissions (#/person/6 months)	0.97	0.07	0.92	0.06	0.053	NA
All-cause inpatient admissions (#/person/6 months)	2.73	0.50	2.64	0.09	0.051	0.148[k]
Outpatient ED visits (#/person/6 months)	3.21	1.22	3.32	-0.11	-0.018	0.210[l]
Medicare Part A and B spending ($/person/6 months)	41,784	10,564	40,458	1,326	0.044	5,160[m]
Service use and spending 12 months before enrollment						
30-day unplanned hospital readmissions (#/person/year)	1.37	0.10	1.33	0.05	0.030	NA
All-cause inpatient admissions (#/person/year)	4.17	0.70	4.11	0.05	0.019	0. 296[k]
Inpatient admissions for ambulatory care-sensitive conditions (#/person/year)	1.25	0.14	1.17	0.08	0.056	NA
Outpatient ED visits (#/person/year)	5.44	1.63	5.84	-0.40	-0.036	0.420[l]
Any primary care services in past year (%)[n]	96.52	92.86	92.55	1.28	0.057	NA

Table II.C.2 *(continued)*

Characteristic	Treatment group (n = 115)	Unmatched comparison group (n = 217,841)	Matched compare-son group (n = 1,000)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS average
Medicare Part A and B spending ($/person/year)	66,703	16,870	65,651	1,062	0.021	10,320[m]
Medicare FFS inpatient spending ($/person/year)	45,491	7,212	42,183	3,308	0.09	NA
Site (%)						
Neighborhood Health Centers of Lehigh Valley	22.6	13.4	21.8	0.8	0.020	NA
Truman Medical Centers	23.5	44.8	25.8	-2.3	-0.052	NA
MultiCultural Medical Group	0.9	4.7	0.7	0.1	0.019	NA
Metro Community Provider Network	53.0	37.1	51.7	1.3	0.027	NA
Omnibus test for balance on matching variables[o]						
p-value		0.9242				n.a.

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code poverty rate merged from the 2012 Five-Year American Community Survey ZIP Code Characteristics.

Notes: Characteristics are measured at the time of enrollment (for the treatment group) or pseudo-enrollment (for the potential and matched comparison groups). The matched comparison group means are weighted based on the number of matched comparisons per treatment beneficiary. For example, if four comparison beneficiaries are matched to one treatment beneficiary, the four comparison beneficiaries each have a matching weight of 0.25.

The chronic condition flags are calculated using one to three years of claims before the enrollment/pseudo-enrollment date (depending on the condition), using the Chronic Condition Warehouse definitions. The flags for Alzheimer's Related Disorders and Senile Dementia used a look-back period beginning three years before enrollment.

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the treatment and matched comparison groups.

[b] The standardized difference is the difference in means between the treatment and comparison groups divided by the standard deviation of the variable, which is pooled across the treatment and comparison groups.

[c] Variables for which we required treatment and comparison members to match on exactly.

[d] Variables that we matched on through a propensity score, which captures the relationship between beneficiaries characteristics and their likelihood of being in the treatment group.

[e] Health Indicators Warehouse (2014a).

[f] Chronic Conditions Warehouse (2014a, Table A1).

[g] DeNavas et al. (2013)

[h] We use 25 of the 27 chronic condition categories defined by the Chronic Conditions Warehouse (see https://www.ccwdata.org/web/guest/condition-categories). We exclude the Alzheimer's Disease and the Acute Myocardial Infarction flags because other flags include these conditions.

[i] The six mental health conditions are conduct disorders and hyperkinetic syndrome, anxiety disorder, bipolar disorder, personality disorders, schizophrenia and other psychotic disorders, and depressive disorders, as defined by the Chronic Conditions Warehouse (see https://www.ccwdata.org/cs/groups/public/documents/document/clin_cond_refer.pdf).

[j] Chronic Conditions Warehouse (2014b, Table B2).

[k] Chronic Conditions Data Warehouse (2014b).

[l] Gerhardt et al. (2014).

[m] Boards of Trustees (2013).

[n] Percentage of beneficiaries with any expenditures for primary care services in the 12 months before enrollment.

[o] Results from an overall chi-squared test indicate the likelihood of observing a set of differences on the matching variables that is as large as what was observed if the treatment and comparison beneficiaries in the matched sample are equivalent on all the matching characteristics indicated. The value of $p = 0.9866$ for the chi-squared test suggests that the two groups are statistically equivalent, because we cannot reject the null hypothesis that they are the same.

Table II.C.2 *(continued)*

*/**/*** Significantly different from zero at the .10/.05/.01 level, two-tailed test.

CHF = congestive heart failure; CKD = chronic kidney disease; CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disease; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = Hierarchical Condition Category; SD = standard deviation.

n.a. = not applicable, NA = not available.

The characteristics of the treatment group are consistent with CSHP's target population: the most frequent users of avoidable hospital care, with eligibility triggered by two or more inpatient hospital or ED visits in the prior six months. The mean HCC risk score of 3.9 is nearly four times the national average (1.0), indicating that the treatment group could be expected to have Medicare spending 3.9 times the national average over the subsequent year. In the 12 months before program enrollment, Medicare spending averaged $66,703, more than six times the national average of $10,320. The mean number of hospitalizations and ED visits (4.17 and 5.44, respectively) were also more than 10 times the national averages. The treatment group members typically had multiple chronic conditions, with an average of 7.8 chronic conditions and 1.4 mental health conditions. A high percentage had congestive heart failure (56.5), chronic kidney disease (62.6), chronic obstructive pulmonary disease (59.1), or diabetes (68.7). These condition-specific rates are each two to five times the national average.

The treatment group differed from the average Medicare population in demographic characteristics and reason for Medicare eligibility. The mean age of the treatment group was 59, and the Medicare FFS average age was 71. It is not surprising then that, for 85 percent of the treatment group, the original reason for Medicare entitlement was disability or ESRD. Just over half of the treatment group was black, compared to 10 percent of the Medicare FFS population. The average poverty rate in the zip codes listed in treatment group members' enrollment data was 27.2 percent, nearly twice the national poverty rate in 2012 of 15 percent.

4. Equivalence of the treatment and comparison groups at the start of the intervention

Demonstrating that the treatment and comparison groups are similar at baseline is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying a contemporaneous differences model—that the outcomes observed for the comparison group during the intervention period are the same on average as the outcomes that would have been observed for the treatment group, had the treatment group not received the intervention.

Table II.C.2 shows that the treatment and matched comparison groups were very similar at the start of the intervention on most matching variables. Means for the treatment group are in the first column, means for the comparison pool are in the second column, and means for the matched comparison group are in the third column. By construction, there were no differences between the treatment and matched comparison groups on the exact matching variable—whether the original reason for Medicare eligibility was either disability or ESRD, as opposed to old age. There were some slight differences between the treatment group beneficiaries and matched comparison group beneficiaries on the variables we matched through propensity scores, but the standardized differences across the propensity score matching variables are all within our target of 0.25 standardized differences. All but one variable, age, was within 0.10 standardized

differences, and most were within 0.05 standardized differences (the 0.25 target is an industry standard; see, for example, Institute of Education Sciences 2014). The omnibus test that the treatment and comparison panels are closely matched on all variables cannot be rejected (p = 0.9866), further supporting that the treatment and comparison groups were similar, based on observable characteristics.

While the groups were well balanced on observable characteristics, it is possible that the groups differ in unobservable characteristics that could affect outcomes and therefore our impact estimates. As described in Table II.A.1, CSHP program staff used information from referrals and EHRs to select beneficiaries that they thought could benefit from the program and exclude those who may be more difficult to treat, including beneficiaries with a history of drug abuse and violent behavior. In addition, beneficiaries had to consent to participate in the program. We could not directly replicate this selection when constructing our comparison group. Thus, it is possible that the treatment group differed from the comparison group on unobservable charactertistics that could be related to outcomes.

The propensity matching technique substantially improved the balance for most variables compared to the full, unmatched comparison pool. This improvement was very important given how different the treatment population was compared to the national Medicare FFS population, as discussed above. Although we did limit the comparison pool to those beneficiaries who had at least one ED visit or hospital discharge over the baseline period, Table II.C.2 shows that the unmatched comparison group was still quite different from the treatment group.

5. Intervention impacts

In this section, we first present sample sizes and unadjusted, mean outcomes, by quarter, for the treatment and comparison groups. These unadjusted outcomes provide context for understanding the impact estimates; however, because they are not regression-adjusted, the differences in mean outcomes are not impact estimates by themselves. Next, we present the results of the primary tests (which are regression-adjusted), by domain. As noted earlier, although we estimate regressions on quarterly outcomes, we implement the primary tests by taking an average of outcomes across the four intervention quarters. We then assess whether primary test results are plausible given the implementation evidence. We end with preliminary conclusions about program impacts in each domain.

a. Sample sizes

In the first intervention quarter (I1), the treatment group includes 115 treatment group beneficiaries and 998 comparison group beneficiaries (see Table II.C.3). This sample is, in general, the same one that we used in matching, as shown in Table II.C.2. However, two matched comparison beneficiaries out of 1,000 are not included in the analysis because they were not observable in updated claims data. (That is, the beneficiaries appeared observable when we first matched them, but in fact they had died or left FFS Medicare. Medicare updated its records to reflect this, and we revised the comparison group accordingly.) The sample decreases in subsequent intervention quarters, as expected, because (1) some beneficiaries did not enroll or

pseudo-enroll early enough to be followed through that quarter, and (2) some treatment or comparison group members exited the sample due to death or otherwise becoming unobservable.

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

Quality-of-care outcomes. The number of ACSCs was lower for the treatment group than the comparison group in I1 and I3, but higher for the treatment group than the comparison group in I2 and I4 (Table II.C.3). In contrast, the number of 30-day unplanned readmissions was consistently lower for the treatment group than the comparison group, with the difference ranging from 4 to 41 percent across the four intervention quarters.

Service use. Across the four intervention quarters, the rates declined slightly from one quarter to the next, but there was no obvious difference between the treatment and comparison group outcomes. The comparison group means were higher than those of the treatment group in some quarters but lower in others.

Spending. Total spending was 5 to 10 percent higher in the treatment group than the comparison group across the intervention quarters, and inpatient spending was 17 to 30 percent higher in the treatment group than the comparison group.

c. Results for primary tests, by domain

Overview. The primary tests are conducted on the quarterly data presented in Table II.C.3. However, the primary tests are based on regression-adjusted estimates, and the impact estimate for each outcome is the average impact over the four intervention quarters. Results of the primary tests differed by domain (Table II.C.4). Tests in the quality-of-care outcomes domain indicate substantively important but not statistically significant favorable effects. Tests in the service use domain were indeterminate. The results for the spending domain indicate substantively important but not statistically significant unfavorable outcomes. Table II.C.4 also shows that, in general, the tests had poor statistical power to detect effects of substantive importance. For example, our tests had a 43.3 percent likelihood of detecting an effect on outpatient ED visits that was, in truth, the size of the substantive threshold. As described above, these results are preliminary because the analyses do not yet cover the full time period that we will include in the final impact analysis in future reports. We also plan to add Medicaid beneficiaries, if the relevant data become available.

Quality-of-care outcomes. The treatment group averaged 206 ACSCs per 1,000 beneficiaries over the four intervention quarters, which was estimated to be 25 fewer admissions (10 percent) than the counterfactual (the estimate of the "counterfactual" is the treatment group mean minus the regression-adjusted contemporaneous differences estimate). However, this difference in admissions was not statistically significant ($p = 0.400$, after adjusting for multiple statistical tests in the domain). The 30-day unplanned readmission rate for the treatment group was 236 per 1,000 beneficiaries, which was estimated to be 98 (29 percent) fewer than the counterfactual, but also not statistically significant ($p = 0.129$, after adjusting for multiple statistical tests in the domain).

Table II.C.3. Sample sizes and unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and comparison groups for Rutgers Center for State Health Policy, by quarter

	Medicare FFS Average	Intervention Quarter 1			Intervention Quarter 2			Intervention Quarter 3			Intervention Quarter 4		
		T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Number of Medicare FFS beneficiaries (unweighted)	49 million	115	998	n.a.	110	891	n.a.	98	664	n.a.	73	510	n.a.
Inpatient admissions for ambulatory care-sensitive conditions (#/1,000 beneficiaries/ quarter)	NA	182.6	214.4	-31.7 (-14.8%)	227.3	208.8	18.4 (8.8%)	112.2	184.8	-72.5 (-39.2%)	301.4	225.2	76.2 (33.8%)
30-day unplanned hospital readmission rate (#/1,000 beneficiaries/ quarter)	NA	287.0	299.2	-12.3 (-4.1%)	227.3	295.0	-67.7 (-22.9%)	224.5	383.3	-158.8 (-41.4%)	205.5	322.6	-117.1 (-36.3%)
All-cause inpatient admissions (#/1,000 beneficiaries/ quarter)	74	860.9	785.6	75.3 (9.6%)	645.5	683.0	-37.6 (-5.5%)	632.7	722.9	-90.3 (-12.5%)	616.4	717.6	-101.2 (-14.1%)
Outpatient ED visit rate (#/1,000 beneficiaries/ quarter)	105	1287.5	1253.7	33.8 (2.7%)	1607.5	1053.0	554.5 (52.7%)	1161.3	1117.8	43.5 (3.9%)	937.6	1442.7	-505.1 (-35.0%)
Medicare Part A and B spending ($/beneficiary/ month)	$860	$5,432	$5,164	$268 (5.2%)	$4,776	$4,342	$434 (10.0%)	$4,839	$4,395	$444 (10.1%)	$4,610	$4,185	$425 (10.1%)
Medicare FFS inpatient spending ($/person/month)	NA	$3,443	$2,927	$516 (17.6%)	$2,919	$2,451	$469 (19.1%)	$3,063	$2,349	$714 (30.4%)	$2,707	$2,214	$493 (22.3%)

Table II.C.3 *(continued)*

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. See Table II.C.2 for sources for the Medicare FFS averages.

Note: The quarters are 3-month periods after a beneficiary's enrollment date (treatment group) or pseudo-enrollment date (comparison group). That is, intervention quarter 1 is the first 3 months after enrollment or pseudo-enrollment, and intervention quarter 2 is months 4 to 6. The means are weighted: each treatment group beneficiary receives a weight of 1; each comparison beneficiary receives a weight equal to the reciprocal of the total number of comparison beneficiaries who match to the same treatment beneficiary. The sample includes beneficiaries whose enrollment or pseudo-enrollment date was between November 1, 2012, and June 30, 2014.

T = Treatment group; C = Comparison group; CMS = Centers for Medicare & Medicaid Services; Diff = difference; ED = emergency department; FFS = fee-for-service.

n.a. = not applicable, NA = not available.

Table II.C.4. Results of primary tests for Rutgers Center for State Health Policy

Primary test definition					Statistical power to detect an effect that is:[b]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual)[a]	The size of the substantive threshold	Twice the size of the substantive thresholdl [a]	Treatment group mean	Regression-adjusted difference between treatment group mean and counterfactual (standard error)[a,c]	Percentage difference [d]	p-value
Quality-of-care outcomes (2)	Inpatient admissions for ambulatory sensitive conditions (#/1,000 beneficiaries/quarter)	Intervention quarters 1 through 4	All observable Medicare FFS beneficiaries attributed to treatment group[f]	-15.0%	28.2	55.1	206	-25 (49)	-10.6%	0.400 [g]
	30-day unplanned readmissions (#/1,000 beneficiaries/quarter)			-15.0%	28.1	54.8	236	-98 (71)	-29.3%	0.129 [g]
	Combined (%)			-15.0%	33.9	67.3	n.a.	n.a.	-20.0%	0.125 [h]
Service use (2)	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)			-15.0%	50.2	90.2	689	-43 (85)	-5.8%	0.426 [g]
	Outpatient ED rate (#/1,000 beneficiaries/quarter)			-15.0%	43.3	82.8	1,248	64 (159)	5.4%	0.550 [g]
	Combined (%)			-15.0%	62.4	97.2	n.a.	n.a.	-0.2%	0.492 [h]
Spending (2)	Medicare Part A and B spending ($/beneficiary/month)			-11.0%	42.1	81.1	$4,914	$512 (448)	11.6%	0.855 [g]
	Medicare FFS inpatient spending($/beneficiary/month)			-15.0%	42.0	81	$3,033	$466 (356)	18.2%	0.889 [g]
	Combined			-13.0%	40.3	78.6	n.a.	n.a.	14.9%	0.883 [h]

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Notes: The results for each outcome are based on a contemporaneous differences regression model that included one to four intervention quarter observations per beneficiary, as described in the text. For each quarter, the model calculates the regression-adjusted difference between outcomes for beneficiaries assigned to the treatment and comparison groups that quarter. The impact estimates from the four intervention quarters were averaged to obtain an average impact estimate across the four quarters. The quarters are 91- or 92-day increments after enrollment in the CSHP program for treatment group members or after the pseudo-enrollment date for comparison beneficiaries. For example, if a treatment beneficiary was enrolled in the CSHP program on July 16, 2013, his or her first intervention quarter is July 16 through October 15, 2013; his or her second intervention quarter is October 16, 2013, through January 15, 2014. The estimates were adjusted for any differences in beneficiary-level covariates (defined in Section II.C.2.d) at the beginning of the intervention period.

Table II.C.4 *(continued)*

The treatment and comparison groups are limited to beneficiaries who were continuously enrolled in FFS Medicare for each of the four quarters before the enrollment or pseudo-enrollment date. Furthermore, in each intervention quarter, the sample consists of Medicare FFS beneficiaries (1) who were enrolled early enough to be potentially followed up for all 91 or 92 days in the quarter, and (2) whose outcomes were observable for at least one day during the quarter. The sample includes those who were in the sample for at least one of the intervention quarters. Outcomes are observable if the beneficiary is alive, enrolled in Medicare Part A and B, not enrolled in a comprehensive managed care plan, and has Medicare as his or her primary payer of medical bills. Outcomes are constructed through December 31, 2014. In each regression model, comparison group beneficiaries are weighted based on the number of matched comparisons per treatment beneficiary. For example, if four comparison beneficiaries are matched to one treatment beneficiary, each of the four comparison beneficiaries has a weight of 0.25.

[a] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted contemporaneous differences estimate.

[b] The power calculation is based on actual standard errors from analysis. For example, in the first row, a 15 percent effect inpatient admissions for ambulatory care sensitive conditions (from the counterfactual of 206 + 25 = 231) would be a change of 35 admissions. Given the standard error of 49 from the regression model, we would only be able to detect a statistically significant result 28.2 percent of the time if the impact was truly 35 admissions, assuming a one-sided statistical test at the $p = 0.10$ significance level.

[d] Percent difference is calculated as the regression-adjusted contemporaneous differences estimate, divided by the estimate of the counterfactual.

[e] These *p*-values test the null hypothesis that the regression-adjusted contemporaneous differences estimate is greater than or equal to zero (a one-sided test).

[f] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

[g] We adjusted the *p*-values from the primary test results for the multiple (two) comparisons made within the domain.

[h] This *p*-value tests the null hypothesis that the contemporaneous differences estimates across the two outcomes in the domain, each expressed as percent change from the comparison group mean, are greater than or equal to zero (a one-sided test).

*/**/*** Significantly different from zero at the .10/.05/.01 level, one-tailed test. No contemporaneous differences estimates were significantly different from zero at any level.

CMS = Centers for Medicare & Medicaid Services; CSHP = Center for State Health Policy; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Awards.

n.a. = not applicable.

The mean percentage difference between treatment and comparison groups across the two outcomes was -20.0 percent (the average of -10.6 percent for inpatient admissions for ambulatory care sensitive conditions and -29.3 percent for unplanned readmissions). This is larger than the pre-specified substantive threshold of 15 percent, but not statistically significant ($p = 0.125$).

Service use. The treatment group averaged 689 all-cause inpatient admissions per 1,000 beneficiaries per quarter over the four quarters following the beneficiary's enrollment date, which was estimated to be 43 admissions fewer than the counterfactual. This 5.8 percent difference between the treatment group mean and the counterfactual was not statistically significant. The rate of outpatient ED visits was 1,248 visits per 1,000 beneficiaries in the four intervention quarters, 64 visits higher than the counterfactual. This 5.4 percent difference was not statistically significant. The mean impact between the two service use outcomes was indeterminate.

Spending. Medicare Part A and B spending for the treatment group averaged $4,914 per beneficiary per month over the first four quarters following the beneficiary's enrollment date, which was estimated to be $512 higher than the counterfactual. This 11.6 percent difference was not statistically significant. Similarly, Medicare inpatient spending for the treatment group averaged $3,033 and was $466 higher than the counterfactual. This 18.2 percent difference was also not statistically significant. The mean percent difference across the two spending outcomes was 14.9 percent, higher than the substantive threshold, but was not statistically significant ($p = 0.883$).

d. Consistency of quantitative estimates with implementation findings

Our quantitative estimates are plausible based on findings from implementation analyses. Although none of our quantitative estimates is statistically significant, substantively favorable impacts on quality of care and substantively unfavorable impacts on spending could reflect implementation findings that program staff tried to get program participants timely and necessary care. For example, treatment group beneficiaries discharged from the hospital might incur longer and more intensive post-acute care services. It is possible that this care reduced readmissions and ambulatory care sensitive admissions by meeting an unmet need, but increased total spending.

e. Conclusions about program impacts, by domain

Based on all evidence currently available, we have drawn the following preliminary conclusions about program impacts in each domain (summarized in Table II.C.5):

- **Quality-of-care outcomes.** The program had a substantively important, but not significant, favorable effect on this domain. Although the mean effect across all tests in the domain was larger in magnitude than our threshold for substantive importance, neither it nor the primary tests for the individual outcomes in the domain (ACSCs and unplanned readmissions) were statistically significant. The lack of statistical significance might be due to the limited statistical power given the small sample sizes.

Table II.C.5. Preliminary conclusions about the impacts of Rutgers Center for State Health Policy, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion	
		Primary test result(s) that supported conclusion	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	Substantively important (but not statistically significant) favorable effect	• The mean effect across all tests in the domain was substantively important but not statistically significant. • None of the individual tests in the domain was statistically significant.	Yes
Service use	Indeterminate effect	• None of the individual tests in the domain were statistically significant or substantively important.	Yes
Spending	Substantively important (but not statistically significant) unfavorable effect	• The mean effect across all tests in the domain was substantively important but not statistically significant. • None of the individual tests in the domain was statistically significant.	Yes

Source: Table II.C.4.

- **Service use.** The program had an indeterminate effect on service use. Neither the mean effect across the two outcomes in the domain (all-cause inpatient admissions and outpatient ED visits) nor the individual tests in the domain were statistically significant or substantively important. This indeterminate effect has two possible interpretations. First, the program may not have an impact on the outcomes for the population and time period covered in this report. Alternatively, the program may have had an effect—and possibly even one that exceeded the substantive thresholds—but, due to the statistical uncertainty in the estimates, we were unable to detect it.
- **Spending.** The program had a substantively important, but not significant, unfavorable effect on this domain. The mean effect across the two outcomes (Medicare Part A and B spending and Medicare FFS inpatient spending) was large, but neither it nor the individual tests in the domain were statistically significant. This finding could reflect the high clinical needs of the patients, reflected in Table II.C.2. and the care team's success in getting patients this needed care.

One should interpret the impact estimates with the beneficiary selection criteria in mind. As noted, because program staff used information available to them from referrals and EHRs, but unmeasurable by us in claims, to select beneficiaries that they thought could benefit from the program, there could be unobservable differences between the treatment and comparison groups that could potentially bias the the estimates. Despite this limitation, and given the evidence from our program implementation analyses, we believe the preliminary conclusions in this report represent the best estimate of the CSHP program's impacts at present.

As mentioned earlier, these conclusions are preliminary because the analyses do not yet cover the full time period that we will include in the final impact analysis in future reports. We also plan to add Medicaid beneficiaries to the final analysis, if the necessary data are available.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

CSHP received HCIA funding to implement a community-based care management program at four provider organizations. The CSHP program used multidisciplinary care teams to connect participants who are frequent users of hospital services to appropriate clinical and social services, help them manage their conditions and overcome socioeconomic obstacles to care. The program aimed to decrease unnecessary hospital admissions and participants' use of emergency department (ED) visits, improve health outcomes, and reduce the average annual cost of care. The four participating sites, which developed site-specific service delivery protocols within CSHP guidelines, implemented the program on schedule. At the two sites visited in April 2015, implementation was facilitated by the ability of front-line staff to flexibly apply the model to meet the needs of individual participants, as well as by staff engagement, stakeholder engagement, and program commitment among leaders in host institutions. Implementation was hindered by frequent changes to staff team roles over the HCIA-funded period, due in part to staff turnover; fragmented supervisory structures; and complexity of patients' needs, combined with gaps in the social services safety net (such as affordable housing). The HCIA Primary Care Redesign Clinician Survey found that most clinicians ranked as extremely important a set of primary care redesign goals that are consistent with the goals of the CSHP program, although few were familiar with the HCIA-funded program itself. Most of the clinicians who responded to the survey practiced at the CSHP implementation sites, but were not HCIA-funded nor members of the CSHP program's care teams.

The impact evaluation found substantively important but not statistically significant favorable effects of the program on quality-of-care outcomes (hospitalizations for ACSCs or 30-day readmissions), no measurable effects on service use (all-cause hospitalizations or outpatient ED visits), and substantively important but not statistically significant unfavorable effects on spending (Medicare Part A and B spending or Medicare FFS inpatient spending). These effects were estimated among Medicare FFS beneficiaries enrolled in the CSHP program through June 30, 2014, over a period of up to four quarters of the primary test period through December 31, 2014. The statistical tests in each of the three domains were poorly or marginally powered to detect effects, so the lack of statistically significant or measurable effects may be because the program did not have effects or that it did, but, due to modest statistical power, our tests failed to detect them. We may find that the program had measurable and statistically significant impacts in one or more of the three domains when the evaluation examines the full intervention period (June 30, 2015).

Our next steps for this evaluation are to (1) monitor CSHP's program implementation reports through June 30, 2015, and plans for sustaining the program beyond the funding period; (2) evaluate trainee and clinician attitudes and experiences with the program in the third year of the award through administered surveys; (3) extend the impact evaluation to include the full

period of program operations and to include Medicaid beneficiaries, if the necessary data are available; and (4) use the implementation findings to help interpret the impact results.

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table A.1 Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: CMS, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.Chronic Conditions Warehouse. "Table B.1 Medicare Beneficiary Counts for Chronic Conditions 2003 – 2012." Baltimore, MD: CMS, 2014b. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf. Accessed November 19, 2014.

DeNavas-Walt, Carmen, Bernadette D. Proctor, and Jessica C. Smith. *Income, Poverty, and Health Insurance Coverage in the United States: 2012.* U.S. Census Bureau, Current Population Reports, P60-245. Washington, DC: U.S. Government Printing Office, 2013.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no 9, 2014, pp. 1672–1679.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Gilman, Boyd, Sheila Hoag, Lorenzo Moreno, Greg Peterson, Linda Barterian, Laura Blue, Kristin Geonnotti, Tricia Higgins, Mynti Hossain, Lauren Hula, Rosalind Keith, Jennifer Lyons, Brenda Natzke, Brenna Rabel, Rumin Sarwar, Rachel Shapiro, Victoria Peebles, Cara Stepanczuk, KeriAnn Wells, and Joseph Zickafoose. "Evaluation of the Health Care Innovation Awards (HCIAs): Primary Care Redesign Programs. First Annual Report, Volumes I and II." Princeton, NJ: Mathematica Policy Research, August 29, 2014.

Hansen, Ben B. "Full Matching in an Observational Study of Coaching for the SAT." *Journal of the American Statistical Association,* vol. 99, no. 467, 2004, pp. 609–618.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

Rosenbaum, Paul R. "A Characterization of Optimal Designs for Observational Studies." *Journal of the Royal Statistical Society, Series B,* 1991, pp. 597–610.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Sanford Health One Care

March 2016

KeriAnn Wells
Jelena Zurovac
Catherine DesRoches
Boyd Gilman
Greg Peterson
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850
Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Lorenzo Moreno

Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

SANFORD HEALTH

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by Sanford Health under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of Sanford Health's One Care program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A) and then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on patients' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF SANFORD HEALTH

Sanford Health received a three-year, $12.1 million dollar HCIA to implement One Care, a medical home model, in 33 practices in Minnesota, North Dakota, and South Dakota; Table I.1 summarizes key features of the program. The goal of the initiative was to integrate behavioral health care services into primary care, with an emphasis on patient screening, workforce development, and health IT tools. The awardee proposed that a system using a redesigned primary care team—including health coaches and behavioral therapists—and addressing chronic disease using a robust health IT system would lead to an engaged patient population. Activated patients work with their care teams and learn to self-manage and improve their care indices, thereby reducing emergency department (ED) visits and hospitalizations and lowering costs. By the end of the award, Sanford Health aimed to reduce potentially preventable admission rates, ED visit rates for Medicare, Medicaid, and CHIP patients with targeted conditions by 20 percent; total cost of care for Medicare, Medicaid, and CHIP beneficiaries with targeted conditions by approximately 3 percent. Sanford Health aimed also to improve quality-of-care outcomes such as optimal care for asthma, diabetes, and hypertension. Sanford Health's HCIA award ended on June 30, 2015.

Table I.1. Summary of Sanford Health PCR program

Program feature	Sanford Health program
Award amount	$12,142,606
Implementation date	April 1, 2013
Award end date	June 30, 2015
Program description	• Implementing a medical home model at 33 Sanford Health primary care practices to screen and treat chronic and behavioral health conditions • At each practice, establishing and training care teams that include primary care physicians (PCPs), nurse health coaches, behavioral health triage therapists (BHTT), and panel managers/care coordination assistants (CCA) • Using health IT to facilitate the medical home for items such as patient screenings, patient portals, telemonitoring, and chronic care protocols for providers
Innovation components	Patient-centered care, integrated care teams, health IT, new forms of patient reporting, workflow redesign
Intervention focus	Practice
Workforce development	Create new positions; implement standard, universal trainings; change roles and responsibilities of existing staff; foster teamwork and collaboration

Table I.1 *(continued)*

Target population	Chronic conditions; behavioral health conditions
Program setting	Provider-based (primary care practices)
Market area	Multistate
Market location	Urban and rural
Core outcomes	• Reduce preventable admissions and ED visits for patients with targeted conditions by 20 percent • Improve quality-of-care outcomes such as optimal care for asthma, diabetes, and hypertension • Reduce overall costs of care for patients with targeted conditions by 3 percent

Source: Review of Sanford Health program reports, March 2015.

Notes: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health IT systems, and undertaking other operational activities.

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how Sanford Health has adapted it over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environment. Finally, we discuss findings related to program sustainability and spillover. We based our evaluation of Sanford Health's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in June 2014 and April 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

Sanford Health's primary effort, known as One Care, was transforming primary care delivery through creation of patient-centered medical homes with fully integrated behavioral health care services for all patients who receive care from a participating practice (see Table II.A.1). To facilitate transformation to patient-centered, integrated care, the awardee invested in health IT to support workforce development, patient screening, quality measurement, and chronic condition management. One Care incorporated tools such as behavioral health screenings, patient synopses, registries, and chronic condition guidelines into the electronic health record (EHR). The awardee also worked to increase patients' use of MyChart (the patient portal) and introduced remote monitoring devices for patients with obesity and hypertension. Sanford Health also used health IT to support staff training on the new model of care. Sanford Health convened a multidisciplinary team of providers, known as the Clinical Skills Development Team (CSDT), which helped design and pioneer One Care and which developed

Table II.A.1. Key details about program design and adaptation

	Patient-centered care model
Target population	• The program was designed to reach all patients at participating practices regardless of insurance, although health coaches especially target adult patients with asthma, diabetes, heart failure, hypertension, and obesity, and children and adolescents with asthma, attention deficit hyperactivity disorder, diabetes, and obesity. • Behavioral health specialists especially target adults and children with anxiety, depression, and alcohol/substance abuse.
Identification strategy	• Care teams review patients' charts to identify those with target conditions or who might benefit from enhanced care. • All patients routinely complete a behavioral health assessment to identify depression, anxiety, and alcohol or substance abuse. For adults, this is the Behavioral Health Screener 6-Item Questionnaire (BH-6), which the awardee developed drawing from four psychometrically validated instruments. Two items were from the Personal Health Questionnaire (PHQ) for depression, two from the General Anxiety Disorder questionnaire (GAD), one from the Alcohol Use Disorders Identification Test (AUDIT), and one from the Drug Abuse Screening Test (DAST). The awardee developed a separate screening tool for patients ages 12 to 17, which collapsed the AUDIT and DAST questions into three questions from the CRAFFT, an alcohol and substance abuse screening tool for adolescents.
Recruitment/ enrollment strategy	• All participating practice patients are considered indirect participants; there is no direct enrollment process.
Service delivery protocol	• Primary care physicians have new resources available to support patients' care, including new care team members, patients' synopses, and chronic condition guidelines. • Panel managers work with patients' data to inform care plans and population management. • Health coaches offer patients chronic condition management education and support, including telephone and e-message follow-up. • Behavioral health triage therapists (BHTTs) offer patients behavioral health triage, assessment, and short-term counseling. • Physicians, nurses, and BHTTs communicate regularly, sometimes via team huddles, to discuss patients' care plans. • Some patients use remote monitoring blood pressure cuffs or scales, which transmit data directly to the EHR for health coaches to review. • New patients, patients visiting for annual check-ups, and those with red flags such as crying complete a six–question behavioral health assessment. Affirmative responses prompt patients to complete longer assessments. Patients either complete the assessment via MyChart or a licensed practical nurse administers the assessment when assigning the patient to a room. • Some practices administer the Patient Activation Measure to assess patients' activation and inform care plans.
Adaptations	• The awardee transitioned from a top-down to a bottom-up approach, drawing from expertise of local physician champions and core practice teams in September 2013. • In November 2013, the awardee hired an integrator for each of the four regions to act as a liaison from practices to awardee leaders. • Individual practices tailored the sequencing and focus of the intervention to their environment. • The awardee introduced remote blood pressure cuffs and scales in March 2014. • In mid-2014, the awardee piloted an addiction navigator in Sioux Falls, South Dakota, and a community health worker in Bemidji, Minnesota.

Source: Interviews from second site visit, April 2015; document review, March 2015.

Note: CRAFFT refers to the topics of each of the six questions: care, relax, alone, forget, friends, trouble.

curriculum for workforce development. Sanford Health houses a Center for Learning, providing online and in–person training; it helped Sanford Health convene a curriculum development team that translated CSDT curriculum into online learning modules for staff (content of trainings discussed in more detail in Section II.1.c).

b. Target populations and patient identification, risk assessment, recruitment, and enrollment

Sanford Health's intervention integrated new staff into practices to offer health coaching and behavioral health services to patients during primary care visits. As shown in Table II.A.1, new services specifically targeted adult and child patients with certain diagnoses. There was no formal enrollment process; rather, all patients who sought care at one of the participating practices had access to the integrated, patient-centered care services available through One Care, but patients with targeted conditions were more likely to receive enhanced care. All Sanford Health practices also incorporated a new behavioral health assessment tool in 2014 and used the screening to identify patients who could benefit from additional behavioral health services.

c. Intervention staff and workforce development

Sanford Health introduced three new staff positions into One Care practices to accomplish patient-centered, integrated care: (1) behavioral health triage therapists (BHTTs); (2) nurse health coaches; and (3) panel managers. (In Sioux Falls, South Dakota, panel managers were rebranded as care coordination assistants (CCAs) to reflect their Medical Assistant credential.) Sanford Health also piloted an addiction navigator and a community health worker (CHW) in two practices. Table II.A.2 provides an overview of new staff and their job responsibilities.

BHTTs helped patients with behavioral health issues such as depression and suicidal thoughts and health coaches helped patients manage chronic conditions such as asthma and diabetes. BHTTs and health coaches reported working as "teamlets," supporting and learning from one another, often finding some overlap between their roles. BHTT salaries were partially award-funded, whereas health coaches predated the award at most sites and their salaries were not award-funded.

Salaries of panel managers/CCAs (who provided support to the health coaches) were almost fully HCIA-funded. Staff reported that the addition of the panel manager/CCA freed nurses to work at the top of their licensure, and that panel managers/CCAs population management work led to improved quality measures.

In mid-2014, Sanford Health piloted two new positions: an addiction navigator in Sioux Falls and a CHW in Bemidji, Minnesota. Sanford Health staffed both positions with lay people, and both positions were HCIA-funded. As of April 2015, the awardee was still determining the optimal way to integrate these two roles into primary care.

Table II.A.2. Key details about intervention staff

Staff members	Credentials	Staff/team responsibilities	Adaptations?
BHTT	Social workers or licensed professional mental health counselors	• Triage patients with behavioral health issues • Conduct behavioral health assessments • Offer short-term therapy in the primary care setting • Coordinate referrals to behavioral health specialists	No
Health coaches	Registered nurses	• Provide chronic condition management education to patients with targeted conditions • Provide ongoing support to patients via telephone calls and electronic messages	No
Panel managers/CCAs	• Lay people (Fargo, North Dakota; Bemidji and Thief River Falls, Minnesota) • Medical assistants (Sioux Falls)	• Manage patients' data to assist health coaches with previsit planning • Use registries to identify patients with chronic conditions • Proactively reach out to patients to administer screenings or schedule visits	No
Addiction navigator	Lay people	• Offer peer support to patients with alcohol or substance abuse issues • Align patients with appropriate addiction treatment • Support patients following addiction treatment	Yes; originally the addiction navigator was located outside of the practice, but was placed inside the practice to increase team's integration and referrals to the addiction navigator. In May 2015, the awardee was still tailoring this position.
CHW	CHW certification (Bemidji)	• Build relationships with local Native Americans • Serve as a liaison between Native American communities and practices • Convene Better Choices Better Health groups to educate Native Americans about chronic condition management • Conduct nonclinical home visits with patients recently discharged from the ED with no primary care physician	Yes; originally the awardee envisioned a cultural advisor position to work on care teams in practices. Sanford Health's cultural liaison advised that the proposed position would be difficult to integrate due to a lack of understanding of Native American cultures. In May 2015, the awardee was still tailoring this position.

Source: Interviews from second site visit, April 2015; document review, March 2015

The HCIA enabled Sanford Health's CSDT to build an online curriculum to train all staff to the new model of care. Training modules covered (1) cultural mindfulness, (2) motivational interviewing, (3) team-based care, (4) chronic disease management and health promotion, (5) the Patient Activation Measure (PAM) survey, (6) trauma-informed care, (7) mental and behavioral health integration, (8) chronic condition guideline use, and (9) psychiatric medications for children and adolescents. Health coaches and BHTTs received most training, although physicians and other clinical staff also completed trainings. Physicians' trainings were pared down to essentials and all trainings were available online to accommodate busy schedules. In addition to these nine training modules, the health coaches and behavioral health staff participated in a chronic care professional training, an eight-module course including additional training in motivational interviewing for which participants received a formal certification.

d. Service delivery protocols

Sanford Health operationalized behavioral health screening in all of its practices—including practices not participating in the HCIA—using the Behavioral Health Screener 6-Item Questionnaire (BH-6) for adults and a modified screening tool for adolescents (see Table II.A.1 for details). In April 2014, the awardee integrated BH-6 into its EHR system and MyChart, enabling patients to complete the screening electronically before their appointments, either at home or on a tablet in the waiting room. Otherwise, the rooming nurse—usually a licensed practical nurse (LPN)—administered the BH-6 and input results into the EHR.

In addition to the BH-6 assessment tool, practices incorporated the PAM survey to evaluate the level of patients' engagement and education needed to help them set and achieve appropriate goals. In April 2015, Sanford Health began incorporating the PAM survey into its EHR system, and several practices had begun administering the PAM survey to patients and using the results to inform care.

Aside from new assessment tools used with patients, the newly developed and trained care teams also used patients' data to facilitate patient-centered care. For example, many practices created registries that they used to monitor and report on patients with targeted conditions. Panel managers used registries to create population health reports, particularly for monitoring patients with asthma and diabetes. Sites also used data to conduct previsit planning, such as determining which patients were due for a BH-6 or could benefit from a visit with the BHTT or health coach. The awardee also built a patient synopsis into the EHR, so care teams had easy access to important data such as screening results, lab data related to chronic illnesses, and PAM scores.

The CSDT created chronic condition guidelines, which the awardee then incorporated into the EHR to provide physicians with decision support. Sanford Health did not mandate that physicians follow guidelines, but physicians used them as resources to inform treatment.

One Care also funded the use of blood pressure cuffs and scales to transmit patients' vital signs directly into the EHR for remote monitoring. Patients had to register devices and have Internet access for the information exchange to function, but many patients also delivered the information manually via MyChart or a telephone call. Health coaches received the data and

used it to inform care plans and goal setting. These remote devices were introduced early in 2014, nearly two years into the innovation, and were not widely adopted.

2. Implementation effectiveness

This section examines the evidence on implementation effectiveness. We evaluate implementation effectiveness based on program enrollment, timeliness, and selected service- and staff-related measures, relying in part on self-reported information included in the awardee's quarterly self-monitoring and measurement reports.

a. Program enrollment

There was no direct enrollment for Sanford Health One Care patients; team-based, patient centered care was available to all practice patients regardless of insurance. From July 2014 to March 2015, the 33 participating practices served an average of 76,679 patients per quarter, exceeding projections in two of the three quarters.

b. Program time line

Sanford Health met many milestones in accordance with its planned time lines, including hiring and training staff and convening core teams at each practice. Three elements of the intervention took longer than expected: incorporating new screening tools, integrating new staff and processes into Bemidji practices, and implementing remote monitoring devices. Practices in Fargo and Thief River Falls had experience with behavioral health integration before the award, but Sioux Falls practices had not, and they reported some delays staffing BHTTs in the Sioux Falls region. This also delayed rollout of the BH-6 in Sioux Falls, because practices were reluctant to screen patients for behavioral health issues until they had appropriate staff to respond. Integrating BH-6 into the EHR also took longer than expected due to technical challenges; it was completed in the second quarter of 2014. In the Fargo region, the awardee piloted EHR integration of BH-6 in one practice. When that practice had fully implemented BH-6 and addressed early technical challenges, the screening was introduced in all Sanford Health primary care practices, including practices that did not participate in One Care.

The awardee also shifted its focus in Bemidji, which delayed One Care implementation in that region. Originally, the awardee proposed to include cultural advisors on the care teams, particularly to advocate for Sanford Health's Native American patients, a sizeable population in Bemidji. However, One Care's cultural liaison cautioned that the goal of full integration would be difficult to achieve in the current climate due to longstanding trust issues and a lack of understanding of local Native American cultures. Instead, the cultural liaison advised One Care to shift its focus to a foundational approach, raising awareness among staff and patients rather than incorporating cultural advisors directly into patients' care. As of April 2015, staff at Bemidji practices worked to engage Native American patients via the CHW and an inspirational video (discussed in greater detail in Section II.3.b).

Another delay occurred with the remote blood pressure cuffs and scales. The awardee reported that many patients struggled to register the devices and that some cuffs were too small

for patients. Staff also stated that the devices were delivered too long after patients agreed to use them, which negatively affected patients' engagement.

Based on these experiences, several staff and administrators commented on the importance of implementing components in a mindful and manageable sequence, suggesting that practices focus on one to three new processes at a time. During the April 2015 site visit, one practice director said, "I think you learn not to tackle the whole elephant at once. Start with three indicators, like asthma, behavioral screenings, and diabetes. If you do too much, the group gets overwhelmed and it collapses."

c. Service measures

The awardee measured and reported the frequency of encounters by staff type, the results of BH-6 screenings, and trends in MyChart adoption. Table II.A.3 shows the number and percentage of One Care practices' patients who had encounters, by staff type, through March 2015. The awardee reported a distinct count of patients each quarter, so patients are counted once during each quarter in which they had an encounter.

Table II.A.3. Sanford Health One Care self-reported encounters

		All encounters		Chronic condition encounters	
Encounter type	Dates tracked	Number	Percentage	Number	Percentage
Behavioral health provider[a]	January 2013 to March 2015	60,371	3.7	41,853	5.4
Health coaches[b]	January 2013 to December 2014	--	--	31,593	4.5
BHTT	January 2014 to March 2015	3,923	0.4	--	--
Assessments		669	0.1	--	--
Triage		2,036	0.2	--	--
Consultations		1,218	0.1	--	--
Addiction navigator[c]	April 2014 to March 2015	110	0.0	--	--

Source: Awardee Self Measurement and Monitoring Results, March 2015.

Note: Denominators include all patients attributed to the 33 practices for dates shown. The awardee also calculated the number and percentage of patients with chronic conditions attributed to the 33 practices who had encounters with behavioral health providers and health coaches.

[a] Psychologists and psychiatrists at One Care practices.

[b] The awardee described these numbers as "grossly underreporting" true health coach encounters due to challenges extracting and integrating data from multiple EHRs.

[c] The awardee offered addiction navigator services at one practice; the denominator includes patients at all One Care practices.

The awardee also reported administering the BH-6 assessment tool to 46,489 adult patients from April 2014 to April 2015, comprising 46 percent of adult patients with qualifying visits at One Care practices. Qualifying visits include visits with new patients and annual check-up visits.

As of March 2015, the awardee nearly met its identification goals for adults with depression and anxiety, and exceeded its identification goals for adolescents with depression and anxiety. However, the awardee fell far short of its identification goals for alcohol and substance abuse among both adult and adolescent patients (Table II.A.4).

The awardee also observed steadily increasing patient adoption of MyChart. From January 2013 to March 2015, quarterly results for patients accessing the portal increased from 11 to 32 percent, although this is still well below the awardee's goal of 70 percent adoption.

Table II.A.4. Sanford Health One Care self-reported behavioral health condition identification statistics

	Adult			Adolescents		
Behavioral health condition	Number	Percentage	Goal percentage	Number	Percentage	Goal percentage
Depression	100,747	16.0	16.1	3,341	7.1	3.9
Anxiety	75,232	12.0	12.3	3,294	7.0	4.7
Alcohol/Substance Abuse[a]	9,321	0.7	3.8 - 5	171	0.4	4.7

Source: Awardee Self Measurement and Monitoring Results, March 2015.

Note: Denominators include all patients attributed to the 33 practices during quarters with available encounter data.

[a] The awardee screens for alcohol and drug abuse separately for adults and together for adolescents. The goal for drug abuse identification for adults is 3.8 percent; the goal for alcohol abuse identification for adults is 5.0 percent; the goal for drug and alcohol abuse identification for adolescents is 4.7 percent.

d. Staffing measures

As of December 2014, One Care had 47 health coaches, 21 BHTTs, 14 panel managers/CCAs, one community health worker, and one addiction navigator deployed in its practices. There were also 25 behavioral health providers embedded in One Care practices. In total, the awardee hired 56.6 new full-time equivalent staff, exceeding its target by 13 percent. The awardee also reported training 6,813 staff cumulatively through December 2014, exceeding its goal of 1,600 trainees. Trainees were counted once for each training they attended; the awardee did not report the number of distinct people trained. However, the awardee did report that as of December 2014, 105 staff had passed the Chronic Care Professional certification exam, including 33 health coaches and 12 BHTTs.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.5 summarizes the major facilitators and barriers to Sanford Health's implementation effectiveness in each domain.

Table II.A.5. Main facilitators and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Perceived relative advantage • Practice-level flexibility	None identified
Implementation process	• Staff engagement • Patient engagement	None identified
Internal factors	• Team dedication • Prior history	None identified
External environment	None identified	• Payment models • Cultural attitudes about alcohol and substance use

Source: Interviews from second site visit, April 2015; document review, March 2015.

a. Program characteristics

Program characteristics include the features of the intervention implemented within an organization, including both core program components and adaptable elements. Two program characteristics especially facilitated implementation of the One Care program: (1) a shared perception that the new model of care was an improvement and (2) flexibility to tailor implementation to the program setting.

First, staff at all levels universally agreed that integrated, team-based care was better for patients and physicians than was traditional volume-based care. Although some physicians were initially reluctant to adopt the new model of care, when they began to see successes among patients who had interacted with nurse health coaches and BHTTs, their attitudes changed. BHTTs reported some difficulty integrating into practices, but noted that their referrals increased as physician buy-in increased. New patient reporting, such as reports showing upward trends in asthmatic patients' asthma control test scores, also facilitated support for the One Care model. Some respondents framed this new use of patients' reports as appealing to physicians' competitiveness, motivating them to improve results relative to their colleagues. Staff at all levels expressed satisfaction with their jobs and unwillingness to return to nonintegrated care.

Sanford Health staff also cited practice-level flexibility as a facilitating factor. For instance, practices were not universally engaging in formal team huddles to discuss incoming patients, although health coaches and BHTTs would always try to touch base with physicians and one another throughout the day. Variations in spatial layout, number of staff, and number of appointments per day influenced teams' ability to formally huddle on a scheduled basis. Program administrators viewed ad hoc huddling as the best approach in settings where formal huddling was infeasible due to space or time constraints. Similarly, although all practices administered the BH-6 and increased behavioral health integration, some practices prioritized certain medical chronic conditions. For example, one practice in Sioux Falls built a diabetes registry and, after it had fully integrated population management for patients with diabetes, began building an asthma registry. A pediatric practice in Fargo prioritized asthma control monitoring, hoping to demonstrate how the panel manager's role in following up with patients and administering the

asthma control test led to an increase in scores, which reportedly occurred. One practice director summarized the notion of practice customization during the April 2015 site visit: "Another takeaway is that you can't cookie cutter it, every group has its own nuances. You need standardization, but you also need flexibility.

b. Implementation process

Two implementation process factors had the greatest effect on the implementation of Sanford Health One Care: staff engagement and patients' engagement.

First, staff engagement was consistently cited as an element of successful implementation. Sanford Health's first step was to create a leadership coalition that could foster both political and financial support within the organization. Awardee leaders also emphasized the importance of engaging both clinical and operational staff, recognizing that operational staff help to move innovation beyond the pilot phase by tailoring job descriptions, priority measures, and integration plans. Leaders quickly learned that a top-down approach would not work and shifted to a model with local physician champions, many of whom served on the CSDT, who could more effectively engage other physicians at their sites. Staff noted that physicians were responsive to data and literature supporting the new model of care. Practices also developed core teams of multidisciplinary stakeholders to drive transformation. Sanford Health designated regional integrators to serve as liaisons between local practices and awardee leadership, supporting the new bottom-up approach to transformation. Practices engaged staff at all levels, including registrars, LPNs, administrative staff, nurse health coaches, BHTTs, and physicians. Regular core team meetings or informal coffee breaks helped build team integration among various staff types and role-specific meetings helped hone skills within individual roles. For instance, during our April 2015 site visit, one BHTT described weekly meetings with other behavioral health staff as "hugely helpful" for brainstorming patients' complex cases.

Second, staff viewed patients' engagement as critical to successfully achieving program goals. Health coaches found motivational interviewing particularly useful for helping patients to set manageable goals. Staff at some practices found PAM very helpful to determine how to best serve a patient based on his or her motivation and knowledge. Many frontline staff described new methods of patient engagement as a paradigm shift, especially for doctors and nurses accustomed to dictating to patients what they must do to be healthy. Teams also reported success with so-called warm handoffs, such as a health coach personally introducing a patient with depression to a BHTT in the practice. Staff attributed this new patient engagement approach to greater perceived success in patients' goal attainment, but emphasized that progress could be slow.

Patients' engagement also included educating them about their conditions and available resources in the community. A practice in Fargo organized patient groups at the Sanford Wellness Center, where patients learned about their illnesses and self-management techniques. Group leaders oriented patients to fitness equipment, a nutritional counselor prepared a healthy snack, and the group engaged in physical activity such as walking or yoga. In Bemidji, where Sanford Health serves a large Native America population, Sanford Health hired the CHW to engage Native Americans in their communities. The awardee also created a motivational video

featuring Billy Mills, a Native American Olympic gold medal winner with diabetes. Mr. Mills spoke at a local Sanford Health-sponsored event, footage of which was included in the video. Although producing this video took longer than expected, the awardee planned to make the video available to patients and providers via hard copy and on its website in the second quarter of 2015.

c. Internal factors

Characteristics of the organization implementing a program and features of the environmental context in which the organization is located can influence implementation. Two internal factors positively affected the implementation of the Sanford Health program: (1) team dedication and (2) a prior history of implementing similar initiatives. First, staff consistently expressed support for team-based, integrated care, emphasizing the value of increased communication and collaboration within clinical teams. Clinical staff and practice administrators worked cohesively to embed new staff into the practices and identified key personality traits suitable for each role before hiring. During site visit interviews, respondents cited persistence and communicativeness as characteristics of successful health coaches and BHTTs. Many staff stated that they were satisfied with their careers and felt that staff satisfaction was high among their peers, especially health coaches and BHTTs. During our April 2015 site visit, one physician noted, "A very delightful result I've seen with [One Care] is development of team unity and staff satisfaction."

Second, practices' past experience transforming care delivery facilitated One Care implementation. Sioux Falls practices had health coaches in place before the award, whereas Fargo and Thief River Falls previously had embedded behavioral health. In both regions, this experience provided the institutional knowledge to integrate new team members and enabled regions to learn from one another. For example, Sanford Health's pre-HCIA work integrating behavioral health in practices in Fargo and Thief River Falls informed behavioral health integration in Sioux Falls. Internal medicine and pediatric practices in the Fargo and Minnesota regions are also certified Minnesota medical homes, and staff at these practices saw One Care as harmonious with their medical home model. In contrast, many respondents in the Fargo region said that the participating family medicine practices were less mature than internal medicine practices in their implementation process, and attributed their delayed progress in part to a lack of prior medical home experience.

d. External environment

Features of an organization's external environment can also influence program implementation. Two external factors presented challenges to the implementation of the Sanford Health One Care initiative: (1) payment models and (2) cultural attitudes about behavioral health. First, Sanford Health's clinician payment model created challenges for sustaining One Care. During our April 2015 site visit, many respondents described payment as having "a foot in each canoe," describing the conflict between Sanford Health's primarily fee-for-service compensation model and its partially risk-based revenue streams. Sanford Health pays physicians based on relative value units−a fee-for-service model−with 5 percent value-based adjustment and pays salaries to other care team members such as health coaches and BHTTs (who also bill payers for

some services). At the same time, Sanford Health also offers a health plan with risk-based contracts and many of Sanford Health's other payers−including Medicaid and Medicare−provide value-based incentives, a growing trend. Physicians and administrators cited challenges transitioning to value-based care under a volume-based payment model. Sustaining new nonbillable services requires Sanford Health to absorb these costs in its operational budget. Sanford Health's status as a large, integrated health delivery system and payer could provide leverage for the organization to adapt to these payment challenges relative to smaller, independent practices. Administrators noted that Sanford Health's sustainability plan would not include use of new Medicare chronic care management billing codes, citing patients' copays, 24-hour EHR access, and burdensome documentation requirements as costs that outweighed the benefits of the new code.

Second, care team members felt that they were successfully identifying and appropriately treating a higher volume of patients with depression and anxiety, but that they had lagged on identifying those with alcohol and substance abuse issues. They attributed this lag to cultural attitudes about alcohol especially, including both patients' and physicians' attitudes. Staff reported that both Dakotas rank very highly in binge drinking, defined as drinking five or more beverages on one occasion. Physicians felt uncomfortable having this conversation with their patients, reporting that most patients who met this definition did not feel that they needed help. Physicians were also reluctant to focus on substance and alcohol abuse unless they felt they had the resources to address the issue with their patients. Hiring the addiction navigator in Sioux Falls was a response to this perceived gap. As of April 2015, the awardee continued to brainstorm ways to more effectively identify and treat patients with alcohol and substance abuse addiction.

4. Sustainability and spillover

Sanford Health leaders indicated that the organization was committed to sustaining and expanding team-based integrated care. Respondents universally agreed that physicians were unwilling to give up the new model after the award, which in turn motivated Sanford Health leaders to find ways to finance new staff's unbillable activities. Sanford Health leaders noted that critical components to sustain were health coaches, BHTTs, panel managers/CCAs, and the BH-6.

Recognizing the difficulty of sustaining the One Care model in a primarily fee-for-service environment, leaders planned to pilot a new physician compensation model in four One Care practices. They expected that the pilot compensation model would pay physicians a salary, possibly enhanced with performance incentives. Although the specifics of the new compensation approach were not finalized, Sanford Health leaders said that the pilot model would hold physicians harmless in the first year, because penalizing them would likely reduce buy-in and morale. Another way that Sanford Health planned to sustain One Care was by leveraging its health plan, securing new risk-based contracts to incentivize improved care management. The awardee also cited Medicaid expansion as an incentive for Sanford Health to focus on improving patients' outcomes without increasing utilization. Sanford Health hoped that these changes would align with patient-centered care and help offset the costs of sustaining nonbillable services.

An interesting spillover effect of One Care was a change in the acuity of patients referred to specialty behavioral health providers. BHTTs believed that integrating new processes into primary care, such as administering the BH-6 and providing in-depth assessments to patients who screened positive, increased referrals to specialty behavioral health practices. Referred patients reportedly were also of higher acuity, thus more appropriate for long-term treatment, to which specialists attributed lower no-show rates. Behavioral health specialists also expressed appreciation that new patients were identified and evaluated sooner, facilitating more efficient and effective treatment in specialty settings. In Sioux Falls, Sanford Health built an increased-capacity behavioral health center, evidence of Sanford Health's commitment to improving behavioral health identification and treatment.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected practices or satellite offices provided important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff are limited to a small number of practice locations and might not reflect the perspectives of clinicians practicing at other sites. In order to assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in fall 2014, the third year of the HCIA-funded program. Data from the survey provide additional insights into the implementation process and experience, as well as the contextual factors that might affect implementation effectiveness at Sanford Health.

In this section, we report on Sanford Health clinicians' views of their daily work life and practice. First, we focus on the contextual factors that can affect program implementation, including the characteristics of the practice's location, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care team functions. We then present data on the alignment of Sanford Health clinicians' views and experiences with the overall goals of the HCIA-funded innovation, as well as their awareness of and participation in the One Care program and their view of the barriers to and facilitators of successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

A total of 123 Sanford Health clinicians responded to the survey, resulting in a response rate of 67 percent. The number of clinicians in each response category do not always sum to 123 due to survey item nonresponse and clinicians who reported that a given question did not apply to their practice and thus did not provide a response. The distribution of clinicians by type, clinical locations, and compensation model is shown in Table II.B.1.

Table II.B.1. Types of clinicians, practices, and compensation sources

Survey item	Number of respondents	Percentage of respondents
Type of clinician		
Physician	99	80%
Nurse practitioner and physician assistant	22	18%
Type of practice		
Group practice (3 or more clinicians)	86	70%
Federally Qualified or other community health center	17	14%
Hospital run by a private for-profit or nonprofit organization	11	9%
Primary compensation source		
Salary adjusted for performance	77	63%
Fixed salary	18	15%
Other (fee for service, time based payment, other)	23	19%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

HMO = health maintenance organization.

Sanford Health clinicians reported working in settings that are advanced in terms of health IT. Nationally, slightly more than half of physicians practice in settings with functional EHRs (Furukawa 2014), but most Sanford Health clinicians reported using health IT at their practice locations. Table II.B.2 shows that most Sanford Health clinicians used EHRs for various functionalities, including use of electronic tracking systems and patient registries, advanced functions that are not in widespread use nationally (DesRoches, Painter, and Jah 2014). Sanford Health also offers patient-facing technologies such as electronic prescription refill and appointment requests.

Table II.B.2. Electronic capabilities for clinicians and patients

Survey item	Number of respondents	Percentage of respondents
Physicians using EHR to		
Receive drug dosing and interaction alerts	121	98%
Access laboratory results	121	98%
Prescribe medications	120	98%
Order tests and procedures	120	98%
Enter clinical notes	120	98%
Track electronic referrals	104	85%
Access patient registries	97	79%
Patients can		
Refill prescriptions	119	97%
Request appointments	117	95%
Email clinician about a medical question or concern	117	95%

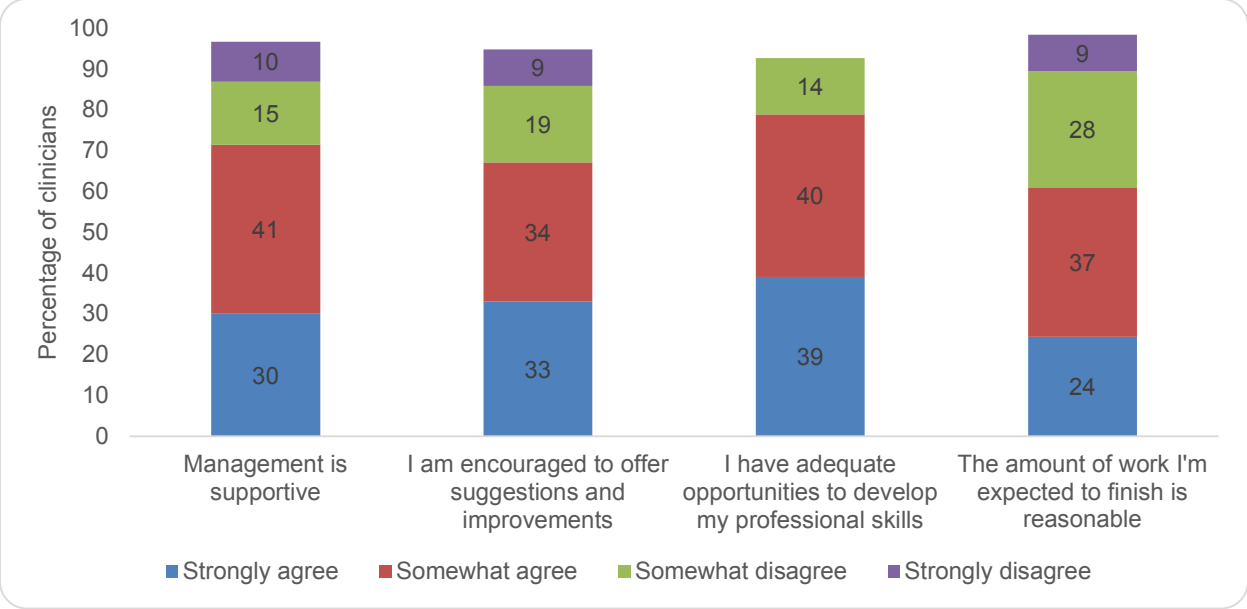
Source: HCIA Primary Care Redesign Clinician Survey, 2014.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall career, level of burnout, and perceptions of their practice environments can all have an effect on the success of program implementation and organizational change. Sanford Health clinicians are generally satisfied with their careers in medicine; 39 percent reported being very satisfied and 37 percent reported being somewhat satisfied (data not shown). Slightly more than one-third experienced some symptoms of burnout at the time of the survey.

Sanford Health clinicians gave similar ratings to their workplace management. As shown in Figure II.B.1, most responding clinicians either somewhat or strongly agreed that their management team was supportive, that they were encouraged to offer suggestions and improvement, that they had adequate opportunities for professional development, and that the amount of work they were expected to complete each day was reasonable.

Figure II.B.1. Workplace ratings



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Totals may not add to 100% due to survey item non-response.

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. As shown in Table II.B.3, one-quarter of responding clinicians strongly agreed and 44 percent somewhat agreed with the statement "It is possible to provide high quality care to all of my patients." Among the major barriers to providing optimal care reported by the majority of Sanford Health clinicians were lack of time to spend with patients—reported by 85 percent of respondents—and insufficient reimbursement, excessive EHR alerts, and patients' inability to pay for care.

Table II.B.3. Perceptions of ability to provide high quality care

Survey item	Number of respondents	Percentage of respondents
It is possible to provide high quality care to all of my patients		
Strongly agree	31	25%
Somewhat agree	54	44%
Neither agree nor disagree	11	9%
Somewhat disagree *or* Strongly disagree	26	21%
Percentage reporting each of the following at least somewhat limits their ability to provide optimal, patient-centered care		
I do not have enough time to spend with patients during visits	105	85%
The level of reimbursement is not adequate	97	79%
I receive too many reminders from my EHR	80	65%
My patients have difficulty paying for needed care	78	63%
I lack timely information about the patients I see who have been care for by other physicians	57	46%
It is difficult for me to obtain specialized diagnostic tests or treatments for my patients in a timely manner	53	43%
It is difficult for me to obtain specialist referrals for my patients in a timely manner	35	28%
I lack adequate information from research evidence to guide my clinical decisions	21	17%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

c. Clinicians' perceptions of care team functioning

Eighty-seven percent of Sanford Health clinicians reported working as part of a care team and, overall, their perceptions of how these teams functioned was positive (data not shown). Most Sanford Health clinicians agreed that members of the care team relayed information in a timely manner (76 percent), had sufficient time for patients to ask questions (72 percent), used common terminology when communicating with one another (73 percent), verbally verified information they received from one another (67 percent), and followed a standardized method of sharing information when handing off patients (62 percent).

d. Alignment with goals of PCR

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The inclusion of the extremely category helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are merely important. In Table II.B.4, we present results based on the proportion of clinicians rating each of these goals as extremely important. The views of Sanford Health clinicians generally aligned with the goals of

PCR. At least fifty percent of clinicians rated 8 of the 13 goals as extremely important. However, it is notable that only one-third of Sanford Health clinicians rated "increasing the number of primary care practices functioning as a patient-centered medical home" as extremely important, given that a specific goal of the Sanford Health One Care program is to create patient-centered medical homes.

Table II.B.4. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Improving care coordination for patients with chronic conditions	77	62%
Improving patients' capacity to manage their own care	73	59%
Improving care continuity in primary care	73	59%
Reducing ED visits	71	58%
Improving appropriateness of care	70	57%
Reducing overall health care spending	69	56%
Increasing access to primary care	67	54%
Increasing the use of evidence-based practice in clinical care	66	54%
Reducing hospital readmissions	58	47%
Improving the capability of health care organizations to provide patient-centered care	51	41%
Improving the capability of health care organizations to provide team-based care	51	41%
Increasing the number of primary care practices functioning as a PCMH	41	33%
Increasing use of EHRs and other health IT	26	21%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: PCMH = patient-centered medical home.

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the One Care program was to change the way care is provided. Program administrators believed that clinicians were critical to that process. Understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on patients' outcomes. For example, if clinicians were aware of the program, had received appropriate and effective training, and believed that One Care would have a positive effect on the care they provide, they may have been more likely to feel more invested in the program's success. Alternatively, those who felt more negatively about the program might have been less likely to enthusiastically implement the intervention. In this section, we report on Sanford Health clinicians' experiences with and perceptions of the One Care program.

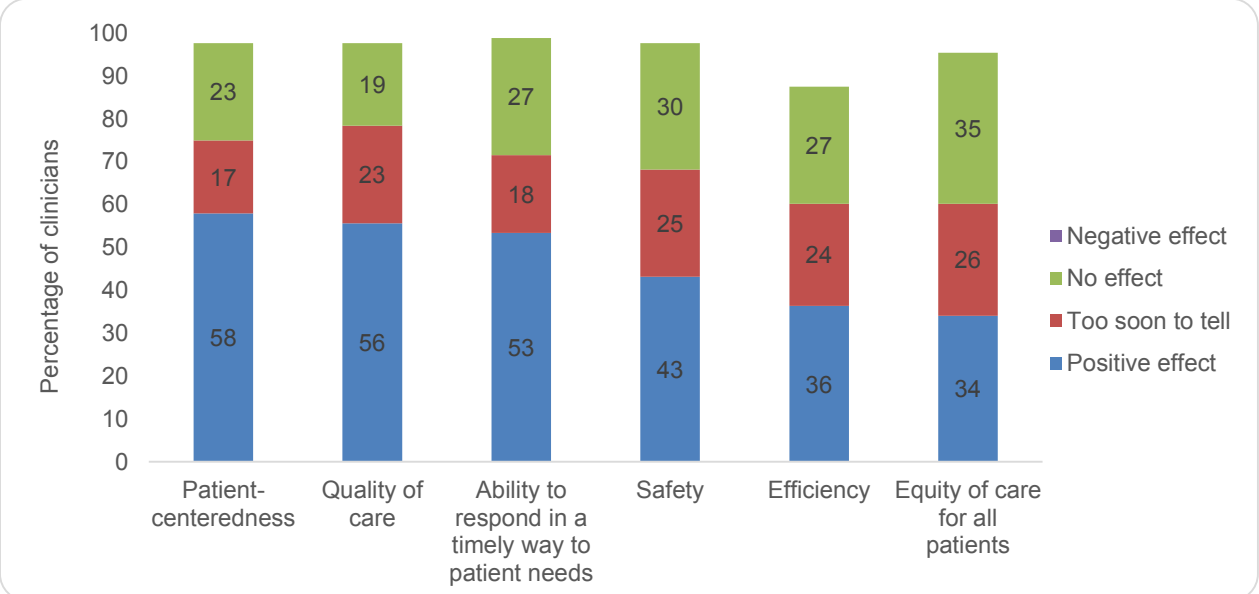
a. Awareness of the program and receipt of training

Almost three-fourths (72 percent) of the Sanford Health clinicians we surveyed were at least somewhat familiar with the One Care program. Of these clinicians, 79 percent had received training related to the program. On average, clinicians received 9.5 hours of program-related training.

b. Perceived effect of program on patients' care

Clinician's perceptions of the effect of the One Care program on the care they provided to patients were mixed. Clinicians were asked about the perceived effect of the One Care program and the barriers to and facilitators of implementation only if they reported being at least somewhat familiar with the program. As shown in Figure II.B.2, about sixty percent of the clinicians who were familiar with One Care believed the program had a positive effect on the quality and patient-centeredness of the care they provided, as well as on their ability to respond to patients' needs in a timely way. Fewer than half of the physicians familiar with One Care believed the program would have a positive effect on equity, efficiency, and safety. Few clinicians perceived an actual negative impact of the program; rather, they believed the intervention would have no effect on the care they provide or that it was simply too soon to tell. However, several clinicians did perceive a negative effect of the program on efficiency of care (data not shown).

Figure II.B.2. Perceptions of effects of program on patients' care



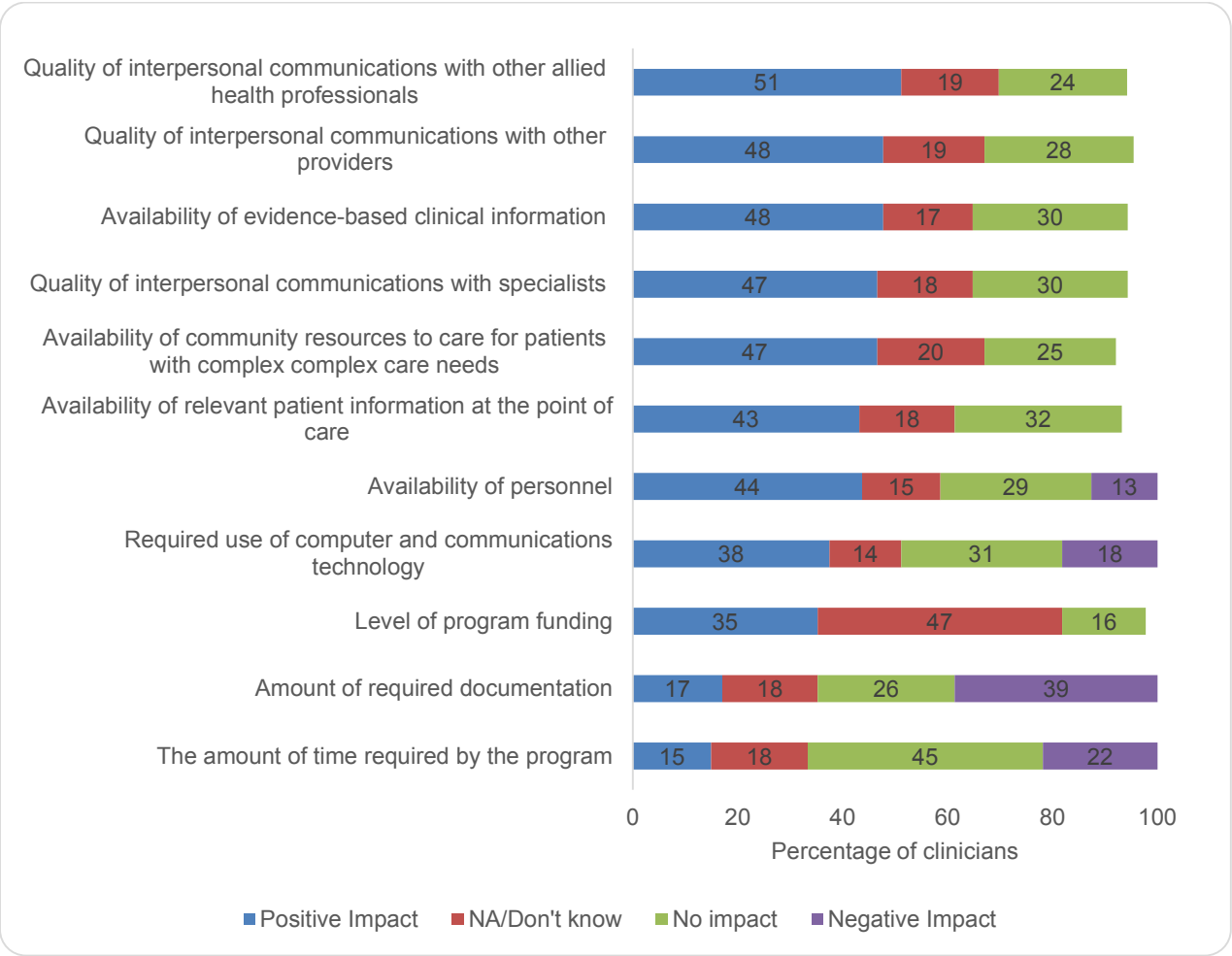
Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Figures are based on the total number of Sanford Health clinicians reporting they were at least somewhat familiar with the One Care program.

c. Barriers to and facilitators of program implementation

Finally, we asked Sanford Health clinicians who were at least somewhat familiar with One Care to rate the effect of a series of barriers to and facilitators of program implementation. The quality of interpersonal communication with other allied health professionals, specialists, and other providers were all seen as having a positive effect on implementation by about half of the surveyed clinicians (Figure II.B.3). The other factor that clinicians perceived as having a positive effect was the availability of evidence-based clinical information. The most often-cited barrier to program implementation was the amount of required documentation, cited by 39 percent of respondents. No other barriers were cited by more than 22 percent of responding clinicians.

Figure II.B.3. Barriers to and facilitators of program implementation



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Figures are based on the number of clinicians who reported being at least somewhat familiar with the One Care program.

4. Conclusions about clinicians' attitudes and behavior

Sanford clinicians surveyed reported higher than average EHR capabilities to support One Care. The majority of clinicians reported satisfaction with their overall career and workplace, although about a third were experiencing symptoms of burnout. The majority of physicians agreed that they could provide high-quality care, but cited barriers to care such as insufficient time, insufficient reimbursement, and excessive EHR alerts. A majority of clinicians reported working in well-functioning care teams. Collectively, the three most important PCR goals to clinicians were (1) improving care coordination for patients with chronic conditions, (2) improving patients' capacity to manage their own care, and (3) improving continuity in primary care. Most clinicians were familiar with One Care, and saw the award as positively affecting patient-centeredness, quality of care, and timeliness of care. Clinicians were less sure if One Care was positively affecting efficiency and equity of patients' care. Most clinicians who were familiar with the program felt that high-quality interpersonal communication with other allied health professionals facilitated program implementation, while the amount of required documentation acted as barrier to implementation.

C. Impacts on patients' outcomes

1. Introduction

In this part of the report, we draw preliminary conclusions, based on available evidence, about the impact of Sanford Health's HCIA program (named One Care) on fee-for-service (FFS) adult Medicare beneficiaries' outcomes in three domains: quality-of-care, service use, and spending. Treatment practices joined the program at different times and, to simplify our analysis, we group them into two cohorts: (1) 21 practices that joined the program from April 1, 2013, to December 31, 2013 (cohort one); and (2) 12 practices that joined the program from January 1, 2014, to December 31, 2014 (cohort two).

Even though Sanford Health's program aimed to reach all patients served by the participating practices regardless of payer, our impact evaluation focuses on Medicare FFS beneficiaries. Lags in Medicaid data availability prevent us from conducting primary tests of effectiveness on the Medicaid population. In addition to excluding the adult Medicaid population, we also excluded six cohort one practices that only serve a pediatric population.We have no plans to analyze impacts on patients served by commercial insurance, because this is not a goal of the HCIA evaluation. As a result, the impact results might not be generalizable to the full population that Sanford's program serves.

In this report, we estimated impacts for the 15 cohort one practices that serve adults. We first describe the methods for estimating impacts (Section II.C.2) and then the characteristics of the 15 participating practices (Section II.C.3). We next demonstrate that these 15 practices were similar at the start of the intervention to the 61 practices we selected as a comparison group (Section II.C.4). Similarity is essential for limiting potential bias in impact estimates. In Section II.C.5, we describe the quantitative impact results, their plausibility given implementation findings, and conclusions about program impacts in each domain. Conclusions in this report are preliminary because the analyses do not include (1) cohort two practices (because they did not join the program early enough to conduct primary tests of effectiveness); (2) the full primary test

period; (3) participating pediatric practices; and (4) Medicaid beneficiaries. The final analysis will include the first two components and potentially the latter two, if Medicaid data become available in time.

2. Methods

a. Overview

We estimated program impacts as the difference in outcomes for Medicare FFS beneficiaries assigned to the 15 cohort one treatment practices serving adults and outcomes for beneficiaries assigned to 61 matched comparison practices that also serve adults, adjusting for any differences between the groups before Sanford Health's HCIA intervention began. To focus the analyses, we specified a limited number of primary tests before examining impact results. Each primary test defined an outcome, population, time period, direction of expected effects, and thresholds regarded as substantively important. We provided the awardee and CMMI an opportunity to comment on the primary tests, and received feedback from CMMI, but not the awardee. We drew conclusions about impacts in each domain based on the results of the primary tests and the plausibility of the primary test results with the implementation findings and secondary quantitative tests (robustness and model checks).

For this report, we analyzed outcomes for Medicare FFS beneficiaries attributed only to cohort one practices because not enough data were available to conduct the primary tests for cohort two practices, given their later start dates. We excluded five pediatric practices from the cohort one analysis because the number of Medicare beneficiaries attributed to pediatric practices would be too small to produce credible impact estimates. We excluded another practice because it was newly opened and thus did not have any Medicare beneficiaries attributed before the intervention began (and very few thereafter).

b. Treatment group definition

The treatment group consists of Medicare FFS beneficiaries assigned to 15 cohort one treatment practices in four quarters before the intervention began (April 1, 2012, to March 31, 2013) and seven intervention quarters (April 1, 2013, to December 31, 2014). We constructed the treatment group in four steps.

First, we attributed beneficiaries to practices using the same decision rule that CMMI uses for the Comprehensive Primary Care initiative. Specifically, in each baseline and intervention month, we attributed beneficiaries to the primary care practice whose providers (physicians, nurse practitioners, or physician assistants) provided the plurality of primary care services in the past 24 months. If there was a tie, we attributed beneficiaries to the practice they visited most recently. Sanford Health provided data on which providers worked in the treatment practices and when.

Second, in each period (baseline and intervention), we *assigned* each beneficiary to the first treatment practice he or she was attributed to in the period, and continued to assign him or her to that practice for all quarters in the period. This assignment rule ensures that—during the intervention period—beneficiaries do not exit the treatment group solely because the intervention

succeeded in reducing their service use (including visits at treatment practices). The definition for the baseline period then corresponds to that of the intervention period so that, across the two periods, interpretation of the population changes over time should be comparable.

Third, we limited the analytic population to beneficiaries targeted by Sanford Health's program. Even though program leadership expected to improve care for all patients attributed to the participating practices, the program specifically targeted improvements in services delivered to patients with one of eight chronic health conditions: anxiety, asthma, diabetes, depression, heart failure, hypertension, obesity, and substance abuse. Sanford Health identified patients with these conditions based on whether administrative data included a diagnosis code for at least one of the targeted conditions. In this report, we present results for this group, and refer to its members as *targeted beneficiaries*. In each intervention program, the targeted group consists of beneficiaries with one of these eight chronic conditions.

Fourth, we applied additional restrictions to define the final sample in each quarter. A beneficiary assigned to a treatment practice in a quarter was included in the sample in that quarter if he or she (1) had observable outcomes for at least one day in the quarter; and (2) lived for at least one day in one of the states with participating practices (Minnesota, North Dakota, or South Dakota) or neighboring states (Iowa or Nebraska).

c. Comparison group definition

The comparison group consists of Medicare FFS beneficiaries assigned to 61 matched comparison practices during each quarter in the baseline and intervention periods, using the same rules we applied to the intervention group. We selected comparison practices that were similar to the cohort one treatment practices during the baseline period on factors that can influence patients' outcomes and factors that influence the decision to participate in the program. Further, we identified the group of targeted beneficiaries served at comparison practices using the same rules as for the treatment group. This section describes how we constructed the matched comparison group whereas Section II.C.4 shows the balance we achieved between the two groups on the matching variables.

To select the matched comparison group, we first identified a pool of 529 nonparticipating potential comparison practices located in the three states with participating Sanford Health practices: Minnesota, North Dakota, and South Dakota. We included in the potential comparison group Sanford Health's nonparticipating practices, because they are more likely to be similar to participating practices than non-Sanford Health practices, given that Sanford Health serves a large part of the region and owns many practices. Two concerns arise for allowing Sanford Health's nonparticipating practices to serve as comparisons: (1) even though they are similar to participating practices on observable characteristics, nonparticipating Sanford Health practices differ in that they were not selected for participation in the award—that is, they might differ from participants on unobservable characteristics; and (2) it is possible that Sanford Health's intervention has been extended to nonparticipating practices. If true, both these concerns can contribute to a bias in impact estimates. However, we do not believe that the risk of bias is substantial. Before the HCIA funds were awarded, Sanford Health planned to roll out the intervention to all of its practices and we found no evidence (in implementation analyses or

informal conversations with Sanford Health) that the final set of 33 participants were selected based on motivation or another unobservable characteristic that could affect outcomes. Further, there is no evidence from implementation analyses or other reason to believe that the HCIA-funded intervention was extended to Sanford Health's nonparticipating practices, with the exception of the behavioral health screening via BH-6 that was implemented in Sanford's non-participating practices. We will investigate further the nature of any behavioral health intervention that followed the screening in non-participating practices. This will help to further assess the risk of bias in impact estimates. Given that only 6 of 61 selected comparison practices are Sanford's non-participating practices and that behavioral health screening and treatment is only one part of the HCIA intervention, we believe that the risk of bias in impact estimates is small.

We then narrowed the pool by excluding practices with characteristics not observed among the treatment group: (1) Indian Health Service practices, (2) practices that do not accept Medicaid, (3) practices located in urban areas, (4) practices that are not owned by a hospital or health system or part of a medical group, and (5) practices with very high or very low values for key matching variables such as practice size and service utilization. After we applied these restrictions, 233 potential comparison practices remained available for matching.

From the 233 potential comparison practices, we used propensity-score matching to select 61 comparison practices that were most similar to the 15 cohort one treatment practices on the matching variables. We matched on characteristics of the practices (number of providers, whether the practice is owned by a hospital or health system, meaningful use of EHRs, and participation in the CMMI Multi-Payer Advanced Primary Care Practice demonstration program); characteristics of the Medicare FFS beneficiaries assigned to the practices (average Medicare spending in the past year, percentage of beneficiaries eligible for Medicare and Medicaid, percentage for whom original reason for eligibility for Medicare was disability, and the percentage of attributed beneficiaries who are Native Americans); and characteristics of the Medicare FFS beneficiaries with one of targeted conditions assigned to practices (mean number of ED visits in the past year, mean number of hospital admissions in the past year, and the mean Hierarchical Condition Category [HCC] score).

The propensity score for a given practice is the predicted probability, based on all matching variables, that the practice was selected for treatment (Stuart 2010). The score collapses the matching variables into a single number for each practice that is used to assess how similar practices are to one another. By matching each treatment practice to one or more comparison practices with similar propensity scores, we generated a comparison group that is similar, on average, to the comparison group on the matching variables. The propensity-score matching approach, however, does not ensure that each comparison practice matches exactly to its treatment practice on all matching variables.

We required each treatment practice to match to at least one, but no more than five, comparison practices and that the ratio of comparison to treatment practices be at least 3:1. This matching ratio increases the statistical certainty in the impact estimates (relative to 1:1

matching), because it creates a more stable comparison group against which the treatment group's experiences can be compared.

d. Construction of outcomes and covariates

We used Medicare claims from April 1, 2009, to December 31, 2014, for beneficiaries assigned to cohort one treatment and comparison groups to develop two types of variables: (1) **outcomes,** defined for each person in each baseline or intervention quarter in which the person is a member of the treatment or comparison group; and (2) **covariates** that describe a beneficiary's characteristics at the start of the baseline and intervention periods, and are used in the regression models for estimating impacts to adjust for beneficiaries' characteristics before the period began. We used covariates defined at the start of each period, without updating them each quarter, to avoid controlling in each intervention quarter for previous quarters' program effects, as this would bias the effect estimates away from detecting true impacts. Appendix 1 describes the methods we used to construct these variables.

Outcomes. We calculated five quarter-specific outcomes that we grouped into three domains:

1. Domain: Quality-of-care outcomes
    a. Inpatient admissions (number/quarter) for ambulatory care-sensitive conditions (ACSCs)
    b. Number of inpatient admissions followed by an unplanned readmission within 30 days (number/quarter)
2. Domain: Service use
    a. All-cause inpatient admissions (number/quarter)
    b. Outpatient ED visits (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
    a. Total Medicare Part A and B spending ($/month)

Four of these outcomes—all but admissions for ACSCs—are outcomes that CMMI has specified as core for the evaluations of all HCIA programs. Our definition of the readmission measure, however, differs from CMMI's standard definition. CMMI typically defines readmissions as the proportion of inpatient admissions that end in an unplanned readmission. Instead, we analyze impacts on the *number* of these unplanned readmissions per quarter because this enables us to look at the total impact on readmissions across the treatment group, rather than readmissions contingent on an inpatient admission (because the intervention might affect the number and type of admissions as well).

Covariates. The covariates include (1) whether a beneficiary has each of 22 chronic conditions (heart failure, chronic obstructive pulmonary disease, chronic kidney disease, diabetes, Alzheimer's and related dementia, depression, ischemic heart disease, cancer, asthma, hypertension, atrial fibrillation, stroke, hyperlipidemia, hip fracture, osteoporosis, rheumatoid

arthritis, bipolar disorder, schizophrenia, drug abuse, alcohol abuse, anxiety, and obesity); (2) HCC scores; (3) demographics (age, gender, and race or ethnicity); (4) whether a beneficiary is dually eligible for Medicare and Medicaid; and (5) original reason for Medicare entitlement (old age, disability, or end-stage renal disease). A rich set of covariates is included to increase the precision of the estimates and to account for any remaining differences between the treatment and comparison beneficiaries on characteristics that might affect the measured outcomes.

e. Regression model

We used a regression model to implement the difference-in-differences design for estimating impacts. For each quarter-specific outcome, the model estimates the relationship between the outcome and a series of predictor variables, assuming that each of the predictor variables has a linear (additive) relationship with the outcome. The predictor variables include the beneficiary-level covariates (defined in Section II.C.2.d); whether the beneficiary is assigned to a treatment or comparison practice; an indicator for each practice (which accounts for stable differences among practices in their outcomes over time); indicators for each post-intervention quarter; and an interaction of a beneficiary's treatment status with each post-intervention quarter. The estimated relationship between the interaction term and outcomes in a given quarter is the impact estimate for that quarter. It measures the average difference between outcomes for beneficiaries assigned to the treatment and comparison practices that quarter, subtracting out any differences between these groups during the four baseline quarters. By providing separate impact estimates for each intervention quarter, the model enables the program's impacts to change the longer the practices are enrolled in the program (which is expected to occur). We can also test impacts over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the impact estimates, allowing for statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. The model used robust standard errors to account for clustering of outcomes across quarters for the same beneficiary and a dummy variable for each practice (fixed effects) to implicitly account for clustering of outcomes for beneficiaries assigned to the same practice. Appendix 2 provides details on the regression methods, including descriptions of the weights each beneficiary receives in the model.

f. Primary tests

Table II.C.1 shows the primary tests for Sanford Health, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold regarded as substantively important (expressed as a percentage change from the counterfactual—that is, the outcome the treatment group would have had in the absence of the HCIA-funded intervention). The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for details and for a description of how we selected each test). However, due to limited data availability, we were able to conduct the primary tests in Table II.C.1 only partially—for cohort one practices and for Medicare FFS beneficiaries, during the fifth, sixth, and seventh intervention quarters (I5–I7) rather than for both cohorts, for both Medicare and Medicaid beneficiaries, and over the entire period of program implementation (I1 through I9). We will report results for the full primary

Table II.C.1. Specification of the primary tests for Sanford Health

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population	Substantive threshold (impact as percentage of the counterfactual)[c,d]
Quality-of-care outcomes (2)	Inpatient admissions for ambulatory care-sensitive conditions (#/person/quarter)	Average over intervention quarters 5 through 9 for cohort one practices; average over intervention quarters 5 through 6 for cohort two practices	FFS Medicare and FFS and managed care Medicaid[e] beneficiaries with a targeted condition assigned to treatment practices	-15.0
	30-day unplanned hospital readmission rate (#/person/quarter)			-15.0
Service use (2)	All-cause impatient admissions (#/person/quarter)			-15.0
	Outpatient ED visit rate (#/person/quarter)			-15.0
Spending (2)	Medicare Part A and B FFS spending ($/person/month)			-2.25
	Medicaid FFS spending ($/person/month)		FFS Medicaid beneficiaries with a targeted condition assigned to treatment practices	-2.25

Notes: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

The One Care program specifically targets improvements in services delivered to patients with one of eight chronic health conditions: anxiety, asthma, diabetes, depression, heart failure, hypertension, obesity, and substance abuse.

[a] We adjusted the *p*-values from the primary test results for the multiple comparisons made within each domain, but not across domains.

[b] The regression models controlled for differences between the treatment and comparison groups during the baseline year when estimating program impacts.

[c] For ambulatory care-sensitive conditions, ED visits, and spending, we used as substantive thresholds 75 percent of the goals set by Sanford Health. For other tests, we relied on substantive thresholds based on Peikes et al. (2011). Sanford Health expected sizable impacts starting in the second year of program implementation (see Section II.C.2.f).

[d] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention.

[e] The extent of Medicaid data availability is uncertain. In the event that Medicaid data are available for only a shorter period, we will revise the period of measurement in our primary tests. We expect to have available Medicaid managed care encounter data for only the state of Minnesota.

ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

tests in our third annual report in 2016, adding cohort two practices and additional follow-up, and depending on Medicaid data availability, possibly also for Medicaid beneficiaries.

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** Sanford Health's central goal is to reduce admissions for ACSCs; ED visit rates; and Medicare, Medicaid, and Children's Health Insurance Program (CHIP) spending. It also seeks to improve clinical and intermediate outcomes that are not easily measured in claims, such as quality of life, functional status measures such as severity of targeted mental health conditions, and the number of encounters and screenings. The primary tests focus on outcomes that Sanford Health aims to affect that are also measurable in claims data: admissions for ACSCs, unplanned readmissions, all-cause hospital admissions, ED visits, and Medicare Part A and B spending.
- **Time period.** Sanford Health expected small impacts during the first year and sizeable impacts in the second year of program implementation. For cohort one practices, our primary tests will cover I5 through I9, corresponding to the period from April 1, 2014, until June 31, 2015, when the award ended. In this report, however, we covered only the period ending December 31, 2014, reflecting data availability. For cohort two practices, our primary tests will cover I5 and I6, corresponding to the period from January 1, 2015, until June 30, 2015.
- **Population.** We chose beneficiaries with one of eight targeted conditions (as defined in Section C.2.b) for the primary tests because (1) the program targeted this group and provided more intensive services to this group than to other patients, (2) Sanford Health specified expected impacts for this targeted population, and (3) the statistical power to detect effects is greatest for this group.
- **Direction (sign) of the impact estimate.** The primary tests are testing for a reduction, relative to the counterfactual, for each of the outcome measures.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we have specified thresholds for what we call substantive importance. For the targeted beneficiaries with one of eight chronic conditions, we set a substantively important threshold of 15 percent for ED visits and inpatient admissions for ACSCs, which is 75 percent of Sanford Health's original estimate of 20 percent for these two outcomes. For Medicare and Medicaid spending, we set a substantive threshold of 2.25 percent, or 75 percent of the anticipated 3 percent. In our fifth quarterly report (Peterson et al. 2015), we had set the substantive threshold for reductions in spending fo*r targeted* Medicare beneficiaries at 15 percent, because we understood that Sanford Health's target of 3 percent referred to *all attributed* beneficiaries and that larger effects would be anticipated for targeted beneficiaries. Another review of documentation revealed that Sanford Health set goals for the targeted beneficiaries, which warranted the change in the primary tests.

  Given that Sanford Health informed us that its initial goals to reduce ED visits were vastly overestimated, using 75 percent of the originally estimated 20 percent reduction might not

be appropriate—that is, a smaller threshold might be regarded as substantively important given the nature of the program and the characteristics of the targeted population. However, because Sanford Health did not provide revised estimates or an indication of how much the goals were overestimated, we proceeded with the above-stated thresholds.

g. Secondary tests

We also conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the non-experimental design or random fluctuations in the data. We will have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. Specifically, we estimated impacts on all-cause admissions, ED visits, 30-day readmissions, and inpatient admissions for ACSCs for targeted beneficiaries during the first year after the practices joined the intervention (I1 through I4). Because Sanford Health expected small impacts in the first year and substantial impacts in the second year, the following pattern would be consistent with an effective program: smaller impacts in the first versus the second year of the program. In contrast, if we had found very large differences in outcomes (favorable or unfavorable) in the first year but not in the second year, this could suggest a limitation in the comparison group, not true program impacts.

h. Synthesizing evidence to draw conclusions

Within each domain, we drew one of four conclusions about program effectiveness, based on the primary test results, the results of secondary tests, and the plausibility of those findings given the implementation evidence. These four possible conclusions are as follows: (1) statistically significant favorable effect (the highest level of evidence); (2) substantively important favorable effect; (3) indeterminate effect; and (4) substantively important unfavorable effect. (We cannot conclude that a program has a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program harms.)

Our decision rules for each of the four possible conclusions are described in Appendix 3. In short, we concluded that a program had a statistically significant favorable effect in a domain if (1) at least one primary test result in the domain was favorable and statistically significant, after adjusting the statistical tests to account for multiple tests (if applicable) within a domain; or (2) the average impact estimate across all primary tests in the domain was favorable and statistically significant. In both cases, we had to determine that the primary test results were plausible given the secondary tests and implementation evidence. We concluded that a program had a substantively important favorable effect if the average impact estimate was substantively important but not statistically significant, and if the result was plausible given the secondary tests and implementation evidence. In contrast, if the average impact estimate was unfavorable (opposite the hypothesized direction), larger than the substantive threshold, and unfavorable effects were plausible given the other evidence, we concluded the program had a substantively important unfavorable effect. Finally, if the tests in a domain did not meet any of these criteria, we concluded that the impact in that domain was indeterminate.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the cohort one practices at the start of the intervention (April 1, 2013). The results can be seen in the second column of Table II.C.2. (Table II.C.2 also serves a second purpose—to show the equivalence of the treatment and comparison practices at the start of the intervention—which we describe in Section II.C.4.).

Characteristics of the practices. Because Sanford Health is a health system, all 15 cohort one practices are regarded as being owned by a system. All treatment practices were located in rural areas, with 13 percent located in health professional shortage areas. Nearly all treatment practices (93 percent) had providers receiving payment from the Centers for Medicare & Medicaid Services for meaningful use of EHRs. Treatment practices had on average 10 providers and a vast majority of providers in these practices had primary care as their specialty (84 percent).

Characteristics of the practices' Medicare FFS beneficiaries. The characteristics of all Medicare FFS beneficiaries assigned to the 15 cohort one treatment practices during the baseline period were, overall, similar to the nationwide FFS averages. The HCC risk score for the treatment group was slightly above the national average (1.05 versus 1.00). Patients in the treatment practices had slightly higher hospital admission rates than the national average, and outpatient ED visit rates similar to the national average. The mean 30-day unplanned hospital readmission rate (11 percent) was lower than the national average of 16 percent, as was the mean Medicare Part A and B spending ($753 per month versus $860).

Targeted beneficiaries (those with at least one of eight targeted chronic conditions) had only marginally higher health care needs during the baseline period than all Medicare FFS beneficiaries assigned to treatment practices (Table II.C.2). Their mean HCC risk scores were slightly higher than the mean for all treatment group members (1.17 versus 1.05). Further, they had approximately 15 percent higher all-cause inpatient admissions, 13 percent more outpatient ED visits, and 13 percent higher Medicare Part A and B spending.

4. Equivalence of the treatment and comparison groups at the start of the intervention

Demonstrating that the treatment and comparison groups are similar at the start of the intervention is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying difference-in-differences designs—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment practices not received the intervention.

Table II.C.2 shows that the 15 cohort one treatment practices and the 61 selected comparison practices were similar at the start of the intervention. For all but four matching variables, standardized differences were within our target of 0.25 standardized differences, and most were within 0.15 standardized differences. (The 0.25 target is an industry standard; for example, see Institute of Education Sciences [2014]). The omnibus test that the treatment and comparison practices are perfectly matched on all variables cannot be rejected ($p = 0.62$), further supporting that the treatment and comparison groups were similar at the start of the intervention.

Table II.C.2. Characteristics of cohort one treatment and comparison practices when the intervention began (April 1, 2013)

Characteristics of practices	Treatment practices (N = 15)	Unmatched comparison pool (N = 233)	Matched comparison group (N = 61)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Exact match variables[c]						
Indian Health Service practice (%)	0.00	0.00	0.00	0.0	n.a.	n.a.
Practice accepts Medicaid (%)	100.0	100.0	100.0	0.0	n.a.	n.a.
Located in a rural area (%)	100.0	100.0	100.0	0.0	n.a.	n.a.
Propensity-score matching variables[d]						
Characteristics of the practices overall						
Owned by a hospital or health system (%)	100.0	77.7	85.3	14.7	0.589*	n.a.
Practice is part of a medical group (%)	0.0	22.3	14.7	-14.7	-0.589*	n.a.
MAPCP demonstration participation (%)	46.7	43.8	47.8	-1.1	-0.022	n.a.
Practice size (number of providers)	9.73	12.62	10.32	-0.58	-0.075	n.a.
Meaningful use of EHRs (%)	93.3	94.0	91.8	1.6	0.062	n.a.
Providers in practice with a primary care specialty (%)	83.9	74.7	84.4	-0.5	-0.020	n.a.
Characteristics of practices' location						
In health professional shortage area (%)	13.3	31.8	36.9	-23.6	-0.515*	n.a.
Characteristics of all Medicare FFS beneficiaries assigned to practices during the baseline year (April 1, 2012, through March 31, 2013)						
Number of beneficiaries	1,106	740	960	145	0.212**	n.a.
HCC risk score	1.05	1.05	1.05	0.00	0.031	1.0
All-cause inpatient admissions (#/1,000 patients/quarter)	77.69	78.29	77.27	0.42	0.031	74[f]
Outpatient ED visit rate (#/1,000 patients/quarter)	107.77	137.26	111.75	-3.98	-0.141	105[g]
Medicare Part A and B spending ($/patient/month)	753	772	772	-20	-0.131	860[h]
30-day unplanned hospital readmission rate (%)	10.98	7.33	9.26	1.72	0.242	16.0[i]
30-day unplanned hospital readmission rate (#/1,000 patients/quarter)k	9.84	10.25	10.01	−0.17	−0.039	n.a.
Inpatient admissions for ambulatory care-sensitive conditions (#/1,000/person/quarter)j	12.80	12.94	13.02	−0.22	−0.046	11.8[k]
Disability as original reason for Medicare entitlement (%)	20.0	25.1	19.8	0.2	0.034	16.7[l]
Percentage dually eligible for Medicare and Medicaid	12.9	15.4	12.3	0.5	0.102	22[m]
Age (years)	72.67	71.36	72.88	-0.21	-0.076	71[n]
Female (%)	61.5	58.3	58.5	3.0	0.397	55.3[l]

Table II.C.2 *(continued)*

Characteristics of practices	Treatment practices (N = 15)	Unmatched comparison pool (N = 233)	Matched comparison group (N = 61)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Percentage Native American or Alaska Native (%)	0.8	0.6	0.6	0.2	0.356	n.a.
Characteristics of targeted Medicare FFS patients[o] attributed to practices during the baseline year (April 1, 2012, through March 31, 2013)						
Number of targeted beneficiaries[j]	837	526	713	125	0.233**	n.a.
HCC risk score	1.17	1.20	1.18	0.00	-0.029	1.0
All-cause inpatient admissions (#/1,000 patients/quarter)	89.04	92.43	90.56	-1.53	-0.092	74
Outpatient ED visit rate (#/1,000 patients/quarter)	121.68	159.29	126.03	-4.35	-0.136	105
Medicare Part A and B spending ($/patient/month)	850	889	873	-24	-0.142	860
30-day unplanned hospital readmission rate (#/1,000 patients/quarter)[j]	11.55	12.66	12.18	−0.627	−0.114	n.a
Inpatient admissions for ambulatory care-sensitive conditions (#/1,000/person/quarter)[j]	15.38	16.15	16.02	−0.641	−0.116	11.8
Omnibus test for balance on matching variables[p]						
p-value	0.62					n.a.

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code (whether an urban zip code or health professionals shortage area) was merged from the American Community Survey ZIP Code Characteristics. Data on meaningful use of EHRs were merged from CMS.

Notes: The comparison group means are weighted based on the number of matched comparison practices per treatment practice. For example, if four comparison practices are matched to one treatment practice, each of the four comparison practices has a matching weight of 0.25.

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the matched treatment and comparison groups.

[b] The standardized difference is the difference in means between the matched treatment and comparison groups divided by the standard deviation, which is pooled across the matched treatment and selected comparison groups.

[c] Exact match means that Indian Health Service practices were excluded from our comparison practices; all practices also had to accept Medicaid and be located in a rural area.

[d] Variables that matched through a propensity score, which captures the relationship between a practice's characteristics and its likelihood of being in the treatment group.

[e] Because we were unable to match within the 0.25 standard on several essential variables when requiring all comparison practices to be owned by a health system, we matched on whether a practice is owned by a health system *or* a medical group. The rationale is that medical groups, like health systems, are able to provide resources to practices that are not available to independent practices.

[f] Health Indicators Warehouse (2014b).

[g] Gerhardt et al. (2014).

[h] Boards of Trustees (2013).

[i] Centers for Medicare & Medicaid Services (2014).

[j] These measures were included in the table for descriptive purposes but were not included in the matching model.

Table II.C.2 *(continued)*

[k] This rate is for individuals ages 65 and above (Truven Health Analytics 2015).

[l] Chronic Conditions Data Warehouse (2014, Table A.1).

[m] Health Indicators Warehouse (2014c).

[n] Health Indicators Warehouse (2014a).

[o] Targeted beneficiaries are those with one or more of eight chronic health conditions targeted by the One Care program: anxiety, asthma, diabetes, depression, heart failure, hypertension, obesity, and substance abuse.

[p] Results from an overall chi-squared test indicate the likelihood of observing differences in the matching variables as large as the differences we observed if, in fact, the treatment and comparison populations (from which we drew the samples) were perfectly balanced. The value of $p = 0.62$ for the chi-squared test suggests that the two groups are well balanced, because we cannot reject the null hypothesis that their characteristics are identical

*/**/*** Significantly different from zero at the 0.10/0.05/0.01 levels, respectively, two-tailed test. No differences were significantly different from zero at the 0.01 level.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; EHR = electronic health record; FFS = fee-for-service; HCC = Hierarchical Condition Category; MAPCP = Multi-Payer Advanced Primary Care Practice.

n.a. = not applicable.

Differences for four variables were outside the target of 0.25 standardized differences: the proportions of (1) attributed beneficiaries who are Native American or Alaska Native, (2) practices owned by a health system or a hospital versus a medical group, (3) practices located in a health professional shortage area, and (4) attributed beneficiaries who are female. Of these four variables, we considered the percentage of Native American beneficiaries to be the most important for matching because Sanford Health selected some practices to participate in the program based on their proximity to Native American reservations. In discussion with CMMI, we determined that it is reasonable to accept the imbalance on this variable, because (1) the magnitude of the difference is small (0.2 percentage points) and (2) the percentage of Native American population is very small at both treatment and comparison practices (0.8 versus 0.6 percent, respectively).

For the three other variables outside the 0.25 standard, we could not improve the standardized difference without causing an imbalance (beyond 0.25 standard) for essential matching variables. We initially required all comparison practices to be owned by a health system. However, we were unable to match within the 0.25 standard on several essential variables. For that reason, we changed our strategy and matched on whether a practice is owned by a health system *or* a medical group. The rationale is that medical groups, like health systems, are able to provide resources to practices that are not available to independent practices. Including potential comparisons practices owned by medical groups enabled us to improve balance on several essential variables. With regard to the imbalance on the proportion of attributed beneficiaries who are female, we believe that this characteristic is unlikely to bias our impact estimates because we control for gender directly using patient-level covariates. We also used practice-fixed effects to capture all time-invariant practice characteristics, which controls for the imbalance on the percentage of practice located in health professional shortage areas, because this characteristic can change only marginally over the length of the intervention period.

The propensity score matching technique improved or did not affect the balance for most variables relative to the unmatched comparison pool, and meaningfully worsened the balance for only one variable—the number of practices in a health professional shortage area.

5. Intervention impacts

In this section, we first present sample sizes and mean outcomes, by quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates that follow; however, the differences in mean outcomes are not regression-adjusted and they are not impact estimates. We present the results of the primary tests, by domain, and the secondary tests results, assessing whether the primary test results are plausible given the secondary tests and given the implementation evidence. We end with preliminary conclusions about program impacts in each domain.

a. Sample sizes

In the first baseline quarter (B1), the treatment group included 12,562 beneficiaries assigned to 15 practices and the comparison group included 36,065 beneficiaries assigned to 61 practices (Table II.C.3). These analysis populations—which are limited to targeted beneficiaries—comprise approximately three-quarters of all Medicare FFS beneficiaries assigned to the practices. The sample sizes increased during the first two or three quarters of the baseline and intervention periods, and decreased slowly but steadily thereafter. This means that after two or three quarters, more beneficiaries move out of the sample (due to death, moving from the region, or switching from FFS to managed care) than are added. As expected, the sum of the comparison group members' weights was roughly equal to the size of the treatment group in each baseline quarter.

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

Quality-of-care outcomes. For both treatment and comparison groups, the rates of ambulatory care-sensitive admissions and 30-day unplanned readmissions declined slightly over the intervention period. The differences between the groups fluctuated, without being consistently positive or negative. The 30-day unplanned readmission rates (number per quarter) were much lower in the treatment group in I1, but otherwise exhibited a similar trend over time.

Service use. All-cause admission rates were similar for treatment and comparison during the baseline period. Over time, they declined steadily for both treatment and comparison groups, with somewhat greater fluctuation among the treatment group. ED visit rates were higher for the comparison group during both the baseline and intervention periods; however, neither the treatment nor comparison group exhibited any obvious trends in this measure. Both increases and decreases were observed, moving together for the treatment and comparison groups.

Spending. The mean spending for the comparison group was similar to the treatment group in most quarters, including both baseline and intervention quarters.

Table II.C.3. Sample sizes and unadjusted mean outcomes for targeted Medicare FFS beneficiaries in the treatment and comparison groups for Sanford Health, by quarter

	Number of Medicare FFS beneficiaries (practices)			Inpatient admissions for ambulatory care-sensitive conditions (#/1,000/quarter)			30-day unplanned hospital readmission rate (#/1,000/quarter)			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/month)		
Q	T	C (no wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (April 1, 2012 – March 31, 2013)																		
B1	12,562 (15)	36,065 (61)	12,514	14.6	16.1	-1.6 (-9.7%)	11.0	12.3	-1.3 (-10.4%)	86.3	91.0	-4.7 (-5.2%)	114.3	125.1	-10.8 (-8.6%)	838.0	863.2	-25.2 (-2.9%)
B2	12,642 (15)	36,491 (61)	12,672	12.3	13.1	-0.8 (-5.9%)	10.9	12.1	-1.2 (-9.7%)	87.5	88.8	-1.3 (-1.5%)	122.1	124.8	-2.7 (-2.1%)	825.2	843.6	-18.4 (-2.2%)
B3	12,702 (15)	36,687 (61)	12,774	16.1	16.3	-0.3 (-1.6%)	11.3	10.8	0.5 (4.2%)	85.6	89.5	-3.9 (-4.4%)	117.9	120.5	-2.7 (-2.2%)	842.7	870.0	-27.3 (-3.1%)
B4	12,320 (15)	35,222 (61)	12,266	15.7	18.8	-3.0 (-16.0%)	11.8	12.3	-0.5 (-4.0%)	91.6	92.9	-1.2 (-1.3%)	109.2	118.4	-9.2 (-7.7%)	871.3	864.3	7.0 (0.8%)
Intervention period (April 1, 2013 – December 31, 2014)																		
I1	12,413 (15)	35,483 (61)	12,392	12.6	15.3	-2.7 (-17.9%)	6.9	10.2	-3.2 (-31.8%)	80.1	86.9	-6.8 (-7.8%)	116.9	120.3	-3.4 (-2.8%)	833.0	870.8	-37.8 (-4.3%)
I2	12,490 (15)	35,742 (61)	12,458	13.5	13.1	0.3 (2.5%)	9.9	11.0	-1.1 (-9.7%)	84.9	83.4	1.4 (1.7%)	113.7	132.6	-18.9 (-14.2%)	859.7	846.9	12.8 (1.5%)
I3	12,539 (15)	35,640 (61)	12,388	12.2	14.4	-2.2 (-15.5%)	9.5	10.5	-1.0 (-9.8%)	74.2	81.0	-6.8 (-8.4%)	113.2	120.4	-7.2 (-6.0%)	859.4	864.6	-5.2 (-0.6%)
I4	12,218 (15)	34,640 (61)	11,988	15.8	15.3	0.5 (3.3%)	10.6	10.9	-0.3 (-3.1%)	88.6	81.1	7.4 (9.2%)	112.7	119.5	-6.8 (-5.7%)	873.3	829.0	44.2 (5.3%)
I5	12,118 (15)	34,191 (61)	11,823	13.9	12.5	1.3 (10.7%)	10.2	11.0	-0.8 (-7.6%)	82.0	81.0	1.0 (1.2%)	120.3	127.6	-7.3 (-5.7%)	869.1	890.1	-21.0 (-2.4%)
I6	11,978 (15)	33,634 (61)	11,611	12.6	11.5	1.1 (9.2%)	9.1	9.8	-0.7 (-7.1%)	78.0	78.6	-0.6 (-0.8%)	125.3	139.3	-14.0 (-10.1%)	857.8	905.0	-47.3 (-5.2%)
I7	11,896 (15)	33,219 (61)	11,461	13.7	14.4	-0.7 (-5.1%)	10.3	10.9	-0.6 (-5.7%)	80.8	80.9	-0.2 (-0.2%)	122.1	133.9	-11.8 (-8.8%)	879.1	877.7	1.4 (0.2%)

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The baseline quarters are measured relative to the start of the baseline period on April 1, 2012. For example, the first baseline quarter (B1) runs from April 1, 2012, to June 31, 2012. The intervention quarters are measured relative to the start of the intervention period on April 1, 2013. For example, the first intervention quarter (I1) runs from April 1, 2013, to June 30, 2013. In each period (baseline or intervention), the treatment group each quarter includes all targeted beneficiaries assigned to a treatment practice by the start of the quarter and who met other sample criteria—that is, they were alive, enrolled in FFS Medicare Parts A and B with Medicare as primary payer, and lived for at least one day in one of the states with participating practices (Minnesota, North Dakota, or South Dakota) or neighboring states (Iowa or Nebraska). In each period, the comparison group each quarter includes all targeted beneficiaries assigned to a comparison practice by the start of the quarter and who met the other sample criteria. See text for details.

Table II.C.3 *(continued)*

The outcome means were weighted so that (1) each treatment beneficiary gets a weight of 1; and (2) each comparison beneficiary gets a weight that is the product of two weights: (a) a matching weight, equal to the reciprocal of the total number of comparison practices matched to the same treatment practice as the beneficiary's assigned practice, and (b) a practice size weight, which equals the average number of targeted beneficiaries assigned to the matched treatment practice during the four baseline quarters divided by the average number of targeted beneficiaries assigned to the beneficiary's comparison practice over those quarters. The difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; Diff = difference; ED = emergency department; FFS = fee-for-service; I = intervention; no wgt = unweighted; T= treatment; wgt = weighted.

NA = not available.

n.a. = not applicable.

c. Results for primary tests, by domain

Overview. For cohort one practices, the primary tests are based on the average impact estimate in I5 through I9. For this report, we had data available only for I5 through I7. Thus, the primary tests in this report reflect impacts over only three of the five primary test intervention quarters (I5 through I7). For each of the five outcomes in the three domains, the regression-adjusted differences between the treatment and comparison groups during the two quarters of the primary test period were small, with one exception: the intervention was associated with a 12.7 percent increase in ambulatory care-sensitive admissions among the treatment group (Table II.C.4). No differences were statistically significant or larger than the substantive thresholds in either a favorable or unfavorable direction.

Quality-of-care outcomes. The rate of inpatient admissions for ACSCs for the treatment group during the primary test period was 12.7 percent higher than our estimate of the counterfactual, and the rate of unplanned readmissions was 4.6 percent lower. (Our estimate of the counterfactual is the treatment group mean minus the difference-in-differences estimate.) Neither difference was statistically significant or substantively large. After combining results across the two outcomes in this domain, the outcomes for the treatment group were slightly higher (4.1 percent) than the outcomes for the estimated counterfactual.

The statistical power to detect effects was marginal for inpatient admissions for ACSCs, but poor for 30-day unplanned readmissions. For example, Table II.C.4 indicates that the tests had a 62.6 percent likelihood of detecting an effect on inpatient admissions for ACSCs that was, in truth, the size of the substantive threshold. Power is worse (55.3 percent) for readmissions because of the greater variation in the outcome.

Service use. The treatment group's average admission rate was 1.7 percent higher and the outpatient ED visit rate was 3.3 percent lower than the estimate of the counterfactual. Neither of these differences was statistically significant nor substantively large. After combining results across the two outcomes in this domain, the outcomes for the treatment group were very similar (0.8 percent lower) to the outcomes for the counterfactual. Power to detect effects that were the size of the substantive thresholds for admissions and ED visits was excellent (99.5 and 99.9 percent, respectively).

Spending. The treatment group averaged $869 in Part A and B spending ($/person/month) during the primary test period, which was 1.9 percent (or $17) lower than the estimated counterfactual. Even though this difference was not statistically significant ($p = 0.24$), it was close to the substantive threshold of 2.25 percent. Statistical power to detect an effect the size of the substantive threshold was low (33.9 percent).

Table II.C.4. Results of primary tests for Sanford Health

Primary test definition					Statistical power to detect an effect that is[a]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual[b])	Size of the substantive threshold	Twice the substantive threshold[c]	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual[b] (standard error)	Percentage difference[d]	*p*-value[e]
Quality-of-care outcomes (2)	Inpatient admissions for ambulatory care-sensitive conditions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 5–7 (April 1, 2014, to December 31, 2014)	All observable[f] targeted Medicare FFS beneficiaries attributed to treatment practices	-15.0%	62.6%	97.3%	13.4	1.5 (1.1)	12.7%	0.847[g]
	30-day unplanned readmissions (#/1,000 beneficiaries/quarter)			-15.0%	55.3%	93.9%	9.8	-0.5 (1.1)	-4.6%	0.441[g]
	Combined (%)			-15.0%	69.8%	99.0%	n.a	n.a.	4.1%	0.688[h]
Service use (2)	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)			-15.0%	99.5%	> 99.9%	80.3	1.3 (3.1)	1.7%	0.557[g]
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			-15.0%	99.9%	> 99.9%	122.5	-4.2 (4.3)	-3.3%	0.271[g]
	Combined (%)			-15.0%	> 99.9%	> 99.9%	n.a.	n.a.	-0.8%	0.384[h]
Spending (1)	Medicare Part A and B spending ($/beneficiary/month)			-2.25%	33.9%	67.4%	$869	-17.1 (24.1)	-1.9%	0.239[g]

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] The power calculation is based on actual standard errors from analysis. For example, in the last row, a 2.25 percent effect on Medicare Part A and B spending (from the counterfactual of $869.0 +$17.1 = $886.1) would be a change of $20. Given the standard error of $24.1 from the regression model, we would be able to detect a statistically significant result 33.9 percent of the time if the impact was truly $20, assuming a one-sided statistical test at the $p = 0.10$ significance level.

Table II.C.4 *(continued)*

[b] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[c] We show statistical power to detect a very large effect (twice the size of the substantive threshold) because this provides additional information about the likelihood that we will find effects if the program is indeed effective. If power to detect effects is less than 75 percent even for a very large effect, then the evaluation is extremely poorly powered for that outcome.

[d] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison groups, divided by the adjusted comparison group mean.

[e] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test.

[f] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

[g] We adjusted the *p*-values from the primary test results for the multiple (two) comparisons made within the service use domain, and (separately) for the two comparisons made within the quality-of-care outcomes domain.

[h] This *p*-value tests the null hypothesis that the difference-in-differences estimates across the two outcomes in the domain, each expressed as percentage change from the estimated counterfactual, are greater than or equal to zero (a one-sided test).

ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

d. Results for secondary tests

As shown in Table II.C.5, the differences in all analyzed outcomes (inpatient admissions for ACSCs, 30-day unplanned readmissions, all-cause admissions, ED visits, and spending) for the treatment group and its estimated counterfactual were small and not statistically significant during the secondary test period: the first year of the intervention (I1 through I4). Even though the estimate of the impacts on readmissions was somewhat large in magnitude (nearly 12 percent), it was not statistically significant and might have occurred by chance, especially given no differences in inpatient admissions. These results help support the credibility of the comparison group because we do not see large differences (favorable or unfavorable) during the first year of the intervention, a period during which we and the awardee did not expect large program effects. This increased confidence in the comparison group, in turn, gives us greater confidence in the primary test results.

e. Consistency of quantitative estimates with implementation findings

The impact estimates in the primary tests are plausible given the implementation findings. The primary tests did not find any effects (favorable or unfavorable) during the primary tests period that were statistically significant or substantively important. The implementation evidence shows that the program was active during the primary test period (I5, I6, and I7). Despite some delays, Sanford Health implemented many interventions and met many milestones within the planned schedule. For example, as described in Section II.A.d, it invested in chronic condition management through increased screenings, incorporated decision support tools such as clinical guidelines into the EHR, and engaged and educated patients through the use of the patient activation tools and the patient portal. Workforce development complemented all activities by educating clinical staff through online and in-person trainings, hiring of staff such as health coaches, and instilling team-based and culturally mindful approaches to care. These implementation metrics indicate that the lack of measured effects is not due to the program failing to deliver a meaningful intervention. However, even with a well-implemented intervention, it is possible that the program was not able to change participants' or providers' behaviors in ways that would affect study outcomes during the primary test period covered in this report (9 of the planned 15 quarters).

f. Conclusions about program impacts, by domain

Based on all evidence currently available, we have drawn the preliminary conclusion that the program impact *is indeterminate in each of the three domains*: quality-of-care outcomes, service use, and spending. These conclusions are summarized in Table II.C.6. We reached these conclusions because (1) in each domain, the primary test results were neither statistically significant nor substantively large; and (2) the secondary tests helped to confirm the credibility of the comparison group used in the primary tests, by showing that there were no estimated effects in the first program year—a period when we and the awardee expected little or no effects.

Table II.C.5. Results of secondary tests for Sanford Health

Secondary test definition				Results			
Domain	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between treatment and the estimated counterfactual (standard error)	Percentage difference[a]	*p*-value[b]
Quality of care	Inpatient admissions for ambulatory care-sensitive conditions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 1–4	All observable[c] targeted Medicare FFS beneficiaries attributed to treatment practices	13.5	-0.1 (1.0)	-1.1	0.443
	30-day unplanned readmissions (#/1,000 beneficiaries/quarter)			9.2	-1.2 (1.0)	-11.6	0.113
Service use	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)			81.9	-0.4 (2.8)	-0.5	0.446
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			114.1	-2.6 (3.7)	-2.3	0.236
Spending	Medicare Part A and B spending ($/beneficiary/month)			856.3	5.3 (22)	0.6	0.595

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Note: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

[a] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison groups, divided by the adjusted comparison group mean.

[b] The *p*-values from the secondary test results were *not* adjusted for multiple comparisons within each domain or across domains.

[c] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

ED = emergency department; FFS = fee-for-service.

Table II.C.6. Preliminary conclusions about the impacts of Sanford Health's HCIA program on patients' outcomes, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion		
		Primary test result(s) that supported conclusion	Primary test result(s) plausible given secondary tests?	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	Indeterminate effect	• Neither of the individual tests in the domain was statistically significant nor substantively important • The combined test across both outcomes in the domain was not statistically significant or substantively important	Yes	Yes
Service use	Indeterminate effect	• Same as above	Yes	Yes
Spending	Indeterminate effect	• The single test in the domain was not statistically significant nor substantively important	Yes	Yes

Sources: Tables II.C.4 and II.C.5

These conclusions have different implications depending on the outcome domain. For the service use domain, the statistical power to detect effects at least as large as the substantive threshold was very good. Therefore, although the program might have had a small effect, it likely did not have a substantively large effect for the study population over the period examined. In contrast, for the other two domains (quality of care and spending), the power to detect effects was marginal or poor. Therefore, the lack of measured effects could mean the program (1) did not have substantively large effects in these domains; or (2) it did, but our statistical tests failed to detect them. It is also possible that the program affected outcomes not measured in the primary tests—for example, behavioral health outcomes, which are one of the focal points of Sanford Health's program.

As mentioned previously, these conclusions are preliminary because the analyses do not yet cover the full time period that we will include in the impact analysis in future reports. Sanford Health continued to implement program activities. It is possible that, when we extend the final evaluation to include an extra four quarters of outcomes, the program will have measurable effects in one or more of the domains, particularly given the delays in implementation of some activities. Finally, this report does not include all practices that participated in Sanford Health's program model; their inclusion might reveal additional impacts and, in addition, would provide greater power to detect impacts.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

Sanford Health received HCIA funding to implement integrated, team-based care in 33 clinics with a focus on early identification and treatment of behavioral health conditions. The program aimed to reduce preventable admissions, ED visits, and total cost of care for Medicare, Medicaid, and CHIP beneficiaries. Sanford Health implemented its program largely as intended, with some delays incorporating screening tools and remote monitoring devices and introducing outreach strategies to Native American patients in one region. Factors that facilitated program implementation included engaged and dedicated staff who saw the new model of care as an improvement and who had flexibility in tailoring the intervention to their practices' unique cultures and workflows. Additional facilitating factors included new patient engagement strategies and prior experience integrating similar initiatives, as well as a robust training curriculum and new uses of health IT and data. Factors that hindered implementation included cultural attitudes about alcohol and substance abuse among both patients and providers and the challenge of encouraging value-based care while paying physicians based on volume. The HCIA PCR Clinician Survey found that most clinicians believed the HCIA-funded initiative would have a positive effect on patient-centeredness, quality of care, and their ability to respond in a timely way to patients' needs.

The impact evaluation estimated program impacts for the 15 cohort one treatment practices that serve adults, which joined the program from April 1, 2013, to December 31, 2013. Results indicate no measurable effects of the program on quality-of-care outcomes (30-day readmissions or inpatient admissions for ACSCs); service use (all-cause hospitalizations or outpatient ED visits); or Medicare Part A and B spending for targeted FFS Medicare beneficiaries during the first nine months of the primary test period (months 12 through 21 after the program began). For service use, the statistical tests were well powered to detect effects, so the lack of measured effects is likely because the program truly did not have substantively large effects. In contrast, for the other domains (quality of care and spending), the lack of measured effects might be because the program did not have effects or that it did but, due to modest statistical power, our tests failed to detect them. The program could have measurable impacts in one or more of three domains when the evaluation is extended to cover the full primary test period (months 12 through 27 after the program began).

Our next steps for this evaluation are to (1) monitor Sanford Health's ongoing program implementation reports through June 30, 2015, and plans for sustaining the program beyond the funding period; (2) evaluate trainees' and clinicians' attitudes and experiences with the program in the third year of the award through administered surveys; (3) extend the impact evaluation to include the full period of program operations, cohort two practices, and, if sufficient data become available on time, pediatric practices and Medicaid beneficiaries; and (4) use the implementation findings to help interpret the impact results.

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

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table B.2. Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf. Accessed November 19, 2014.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no. 9, 2014, pp. 1672–1679.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Health Indicators Warehouse. "Medicare Beneficiaries Eligible for Medicaid (percent)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014c. Available at http://www.healthindicators.gov/Indicators/Medicare-beneficiaries-eligible-for-Medicaid-percent_317/Profile/ClassicData. Accessed August 4, 2015.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

Peterson, Greg, Laura Blue, Lorenzo Moreno, Sandi Nelson, Kate Stewart, Frank Yoon, Keith Kranker, Randall Blair, Eric Lammers, David R. Mann, Sean Orzol, Purvi Sevak, Andrea Wysocki, Jelena Zurovac, Thomas Bell, Andrew McGuirk, Ken Peckham, Boyd Gilman, Cara Stepanczuk, and Brenna Rabel. "Evaluation of the Health Care Innovation Awards (HCIAs): Primary Care Redesign Programs. Quarterly Report to CMS: Quarter 5 (October 1, 2014 – December 31, 2014)." Princeton, NJ: Mathematica Policy Research, March 16, 2015.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21.

Truven Health Analytics. "AHRQ Quality Indicators, Prevention Quality Indicators v5.0 Benchmark Data Tables." Prepared for the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. Santa Barbara, CA: Truven Health Analytics, March 2015. Available at http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V50/Version_50_Benchmark_Tables_PQI.pdf. Accessed Aug 18, 2015.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for TransforMED

March 2016

Rosalind Keith
Sean Orzol
Mynti Hossain
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Lorenzo Moreno

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

FINDINGS FOR TRANSFORMED

This individual program report summarizes the findings to date from our evaluation of the primary care resdesign (PCR) program implemented by TransforMED under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the TransforMED program and Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behavior of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the program on patients' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF TRANSFORMED

TransforMED, a national learning and dissemination contractor, received a three-year, $20.8 million award to implement the patient-centered medical neighborhood (PCMN) program. TransforMED provides technical assistance to 14 health systems, which in turn recruited 90 primary care practices across the United States, to implement the PCMN program. Table I.1 summarizes key features of the program. Most participating health systems were part of VHA, a national network of nonprofit health systems. TransforMED refers to each health system and the participating practices as a community. Through PCMN implementation, TransforMED hoped to reduce overall health care costs for Medicare and Medicaid beneficiaries, improve the patients' health and experiences with care, and expand the model to additional primary care practices. TransforMED's HCIA award ended in June 2015.

Table I.1. Summary of TransforMED PCR program

Program feature	TransforMED program
Award amount	$20,750,000
Implementation date	November 2012
Award end date	June 2015
Program description	• Implement population management software in participating primary care practices • Implement cost management reporting in participating primary care practices • Community-wide participation in PCMN collaboration and shared learning activities
Innovation components	Health IT (population management systems and cost management reporting)
Intervention focus	Primary care practices
Workforce development	Establish one health coach in each practice and three super users in each community (individuals selected to facilitate implementation of Cobalt Talon software and reporting)
Target population	All insured patients treated at participating practices
Program setting	Provider-based (primary care practices)
Multistate	Multistate
Market location	Urban, rural

Table I.1 *(continued)*

Core outcomes	• $49.5 million reduction in overall care costs for Medicare and Medicaid beneficiaries • 15 percent improvement in condition-specific quality measures • 25 percent improvement in patients' experiences • Expand program to 18 to 20 additional practices in each community

Source: Review of TransforMED program reports, March 2015.

Note: The implementation date represents when the communities began taking concrete steps toward launching program components by implementing health IT systems and undertaking other operational activities related to the program.

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we provide a detailed description of the PCMN program, highlighting changes in program design over time. Second, we review the evidence on implementation effectiveness, including an assessment of measures of enrollment, the PCMN implementation time line, and other service- and staff-related metrics. Third, we examine the facilitators and barriers that influenced implementation effectiveness, specifically those related to program characteristics, implementation processes, internal factors, and external factors. Finally, we discuss TransforMED's plans for program sustainability and spread. We based our evaluation of TransforMED's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in April 2014 and March through April 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

The TransforMED program provided tools to practices so they could more effectively use data to improve clinical processes, such as systematically identifying patients in need of preventive screenings or who would benefit from a care plan, all in the larger effort of providing better care to patients. Participating practices may also have implemented new patient care protocols as a result of having access to these tools. However, the TransforMED program did not require participating practices to implement changes to their patient care protocols. The program involved the implementation of two primary components: population management systems and cost management reporting functions. Although 13 practices opted not to implement the population management system software, most participating practices used new forms of patients' data generated by these two components to implement practice transformations. After the majority of practices implemented population management systems and cost management reporting functions, TransforMED added a third program component, data analytics, which involved integrating population management and cost management data (Table II.A.1).

Table II.A.1. Key details about program design and adaption

	Program Component		
	Population management systems	Cost management reporting systems	Data analytics
Target population	All insured patients treated by a PCP at a participating practice are eligible; however, practices risk-stratified patient panels to target specific patient populations.		Patients who are treated by a PCP at a participating practice whose cost of care could be reduced through improved care coordination across providers in the PCMN community
Patient identification	To risk-stratify, practices used quality indicators, cost and utilization metrics, anecdotal partcipant information, internal automated algorithms, or the MARA model.		n.a.
Patient recruitment and enrollment	No eligibility criteria were established for patient enrollment into the program. All insured patients were passively enrolled in the program; no formal enrollment procedures existed.		No eligibility criteria were established for patient enrollment into the program. All insured patients were passively enrolled in the program; no formal enrollment procedures existed.
Service delivery protocol	The program did not directly change the way in which providers treated or interacted with patients, but gave them information and data tools to improve their existing proceses.		n.a.
Adaptations	None	Due to technical demands involved in generating customized reports, practices mostly used the standardized reports.	None

Sources: Interviews from second site visit, April 2015; document review, March 2015.

MARA = Milliman Advanced Risk Adjusters; PCP = primary care provider.

n.a. = not applicable.

TransforMED worked with Phytel, a health care technology company, to implement two types of population management software in 78 of the 90 participating primary care practices (71 system-employed and 7 non-system-employed practices) across 14 communities. The Phytel Insight™ software program organized clinical data by patient and population characteristics and quality indicators. The Phytel Coordinate™ software program automated care management processes within practices by providing care teams with the following capabilities: (1) patient attribution, which involved assigning patients to primary care providers who were responsible for coordinating their care needs; (2) risk-stratification, which involved assessing a patient's health risk status and categorizing the patient based on his or her care needs; and (3) patient outreach, which involved targeting communications to patients based on their individual care needs, as measured by quality indicators. TransforMED expected that the combination of these data organization and automated care management capabilities would enable practices to target improvements to specific quality indicators and patient populations—for example, identifying

patients due for a mammogram and sending them automated reminders to schedule the procedure.

Participating primary care practices implemented the Cobalt Talon cost management reporting software to support the analysis of Medicare fee-for-service (FFS) claims and to generate dashboard reports on utilization and cost of care at the community and practice levels. Because the cost management data provided a retrospective review of claims information, TransforMED initially encouraged practices to generate these reports with a three-month lag period. However, even with a three-month lag period, claims processing was incomplete and practices therefore saw large fluctuations in the cost management measures from month to month. Because of the problems with claims reporting in the early stages of PCMN implementation, the practices did not trust the claims-based cost management reporting. TransforMED realized that claims processing was almost 100 percent complete after a six-month lag period and created a rule that 85 to 90 percent of the claims had to be complete before practices generated cost management reports from Cobalt Talon. TransforMED originally expected that practices would customize Cobalt Talon cost management reports to focus on the unique utilization and cost issues within their community. However, because of the investment of time needed to learn how to generate customized reports, practices generally relied on the standardized reports developed by Cobalt Talon.

After most practices had implemented Phytel and Cobalt Talon, TransforMED rolled out the third program component, data analytics, to integrate population management and cost management data and move communities closer to functioning as PCMNs. To implement data analytics, TransforMED met with leadership in most PCMN communities to discuss the integration of Phytel and Cobalt Talon reporting functions and the use of patients' data in new ways to manage their care. The data analytics component involved integrating the quality indicator data reported in Phytel and the cost management data reported in Cobalt Talon to target patients whose care could be improved and whose cost of care could be reduced through improved coordination of care across providers within the PCMN community.

TransforMED introduced two tools to support the data analytics component: patient profile reports and the Cave Grouper Efficiency Measurement Tool. The patient profile reports were a web-based tool that gave participating practices information on all the services received by a patient and a risk score based on cost and utilization metrics. The primary challenge with the reports was that they did not provide real-time information and were designed to look at data from only a historical or planning perspective. Initially, community buy-in for the patient profile reports was lower than TransforMED had anticipated. However, the development of additional capabilities in the web-based patient profile report, such as identifying patients who were potentially eligible for care transitions management codes or Medicare wellness visits, increased community buy-in related to the reports. The Cave Grouper tool offered participating primary care practices the ability to compare physicians' efficiency (based on cost and utilization metrics) against a national peer group, producing efficiency scores for physicians by certain diagnoses and procedures. The Cave Grouper tool also gave individual providers, practices, and communities information on patient referral patterns and provider efficiency related to specific disease conditions or specialties.

The data analytics component revealed to TransforMED that the total cost of care was more highly correlated with specialty care than it was with primary care. This finding led to discussions between TransforMED and community leadership about spreading the PCMN program to specialty care providers in PCMN communities, rather than focusing solely on spreading the PCMN program to other primary care providers in the community. As of July 2015, this adaptation has not been implemented.

b. Target populations, risk-assessment, recruitment, and enrollment

TransforMED did not establish eligibility criteria for patient enrollment into the program; all insured patients treated at participating practices were eligible to receive PCMN-related services. However, with the focus on population management systems and cost management reporting, participating practices began stratifying patient panels based on health care needs. Practices used various indicators of risk to stratify their panels, including quality indicators, cost and utilization metrics, anecdotal patient information, and internally automated algorithms based on various criteria. Several practices relied on the MARA model, which calculated risk scores based on Medicare FFS claims. TransforMED encouraged practices to identify patients with multiple emergency department (ED) visits as high risk. But they did not require that practices focus their PCMN-related services (for example, by systematically identifying and contacting patients who would benefit from a care plan or those in need of preventive screenings) on high-risk patients. Some practices began PCMN implementation by focusing on low-risk patients with care gaps; many existing practice workflows were already designed to support that relatively healthy patient population. Later, as their population management systems and procedures were systemized, practices began to focus care management activities on rising-risk or high-risk patients. Essentially, participating practices used new sources of data to improve care within certain patient populations; across communities and practices, those populations were defined differently.

c. Intervention staff and workforce development

TransforMED employed a variety of strategies to facilitate PCMN implementation and develop a PCMN workforce in the 14 communities (Table II.A.2). TransforMED provided regular on-site and virtual support to practices, health systems, and communities. Over time, TransforMED began to customize the support depending on an individual community's needs. TransforMED continued to use individual system plans and individual practice plans as a framework for monitoring 30-, 60-, and 90-day goals and action planning within each community. In the ninth quarter of PCMN implementation (July through September 2014), TransforMED created a forum for the 14 communities to network within the PCMN program. TransforMED also organized PCMN system networking calls for community leadership to connect and exchange learnings related to patient-centered medical home (PCMH) implementation. Five communities participated in an in-person cross-community learning collaborative to discuss progress and challenges related to successfully establishing a PCMN within their communities. From this collalborative emerged a monthly virtual meeting across communities. The topics of these meetings included group or shared visits, evidence-based health coaching techniques, collaboration with specialists under the PCMN, integrating behavioral health, and managing inappropriate use of the ED.

Table II.A.2. TransforMED technical assistance and training activities

Technical assistance	
Biannual communitywide learning collaboratives	Representatives from the health systems and practices attended these biannual meetings to discuss PCMH concepts (such as patient attribution, risk-stratification, care coordination, and care teams) and best practices for implementing program components.
Monthly conference calls	In the first and second years of the award, TransforMED held monthly conference calls with each health system and made up to four visits to individual practices.
Quarterly community leadership meetings	TransforMED organized on-site quarterly meetings with leadership in each community to review updated data, analytics, trends, and practice and health system-aligned initiatives, and to discuss priorities for the community.
Cross-community learning and PCMN collaboration	Five communities participated in a cross-community learning collaborative to discuss implementation progress and share experiences with overcoming challenges to successful PCMN implementation.
Workforce development	
Cobalt Talon training	TransforMED selected super users (facilitators of Cobalt Talon software implementation) in each community to attend a two-day training conducted by Cobalt Talon in which super users learned how to generate reports from the Cobalt Talon system and discussed health IT, clinical integration, and PCMH and PCMN concepts. TransforMED hosted two follow-up telephone calls with super users to discuss their experiences and challenges using the Cobalt Talon reports.
Clinical health coach training	Each participating practice selected a health coach to attend a clinical health coach training session conducted by the Iowa Chronic Care Consortium. Patients learned about motivational interviewing, evidence-based health coaching, population health and risk-stratification, and coaching using the Myers-Briggs Type Indicator®.

Source: Review of TransforMED program reports, March 2015.

TransforMED anticipated that participating practices would develop and independently fund a nurse care manager role or a comparable position to use the population management information generated by Phytel, although this was not a requirement of program implementation. All 14 communities hired at least one individual to fill a care management-type role and used them in varying capacities. Some focused on coordinating transitions of care, which was an initial focus of PCMN implementation. Some focused on providing self-management support to high-risk patients. After an initial in-person training in Year 2 of the award, referred to as clinical health coach training, TransforMED continued to organize virtual training opportunities for patients. TranforMED and Cobalt Talon identified three individuals in each community to facilitate the implementation of the cost management reporting component. These super users, who included a range of clinical and administrative staff from participating health systems and practices, attended a training in Year 2 of the award to learn how to generate reports from Cobalt Talon (Table II.A.3).

d. Intervention protocols

The TransforMED program provided tools to practices so they could more effectively use data to improve clinical processes, such as systematically identifying patients in need of preventive screenings or who would benefit from a care plan, all in the larger effort of providing better care to patients. The TransforMED program did not specify changes to patient care protocols, although practices could have implemented new patient care protocols as a result of having access to the tools they received through the program.

Table II.A.3. Key details about intervention staff and workforce development

Program component	Staff members	Staff /team responsibilities	Adaptations?
Population management systems	One existing staff member in each practice (for example, nurse care manager or comparable position)	Served as the main point of contact for using the population management information generated by Phytel; however, practices were not required to develop this role; referred to as clinical health coaches	No significant barriers identified
Cost management reporting	Three existing clinical and/or administrative staff in each community	Served as the main point of contact for implementation of the cost management reporting generated by Cobalt Talon; referred to as super users	Relied on standardized reports developed by Cobalt Talon; developing their owncustomized reports required too much staff time to learn to build and run

Source: Interviews from second site visit, April 2015; document review, March 2015.

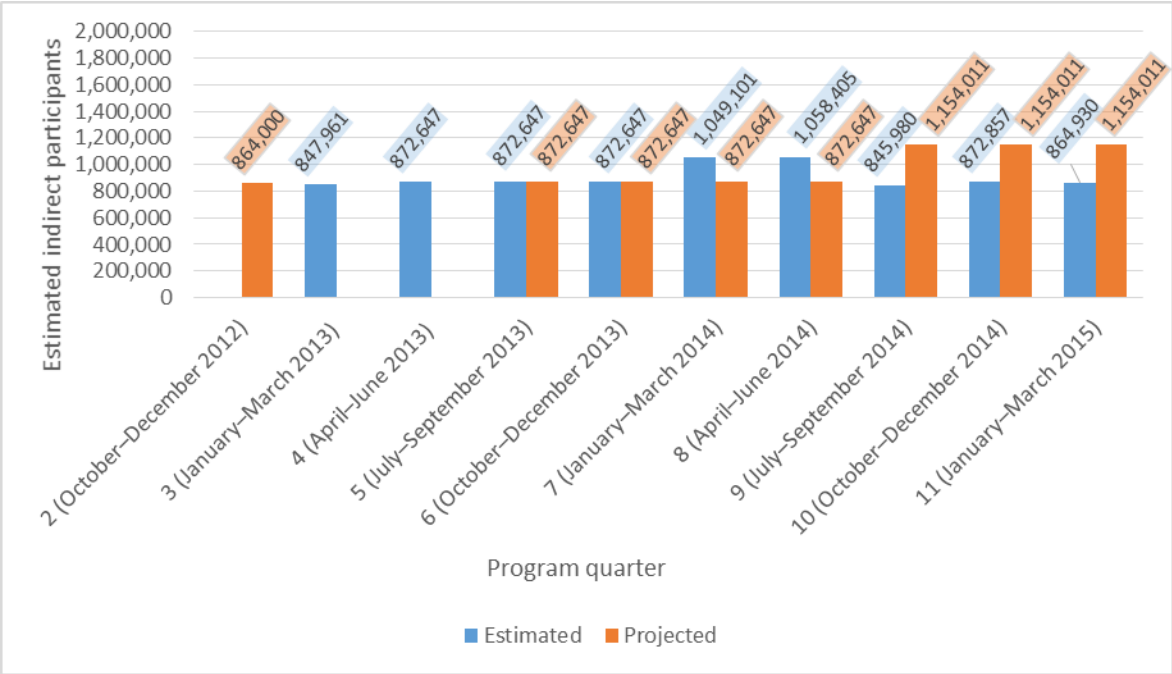
2. Implementation effectiveness

In this section, we examine the evidence that the PCMN program was implemented effectively. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators and self-reported information included in TransforMED's quarterly self-monitoring and measurement reports.

a. Program enrollment

TransforMED successfully met its goals of enrolling 15 health systems and 90 practices to participate in the program. Although TransforMED did not establish eligibility criteria for patient enrollment in the program, it expected 864,000 indirect program participants—patients insured by Medicaid, Medicare, or a commercial payer and treated by a primary care provider at a participating practice—in the 1st quarter of the program. The program started collecting program participant counts during the 2nd quarter (October through December 2012), but enrolled participants starting in the 3rd quarter (January through March 2013). The number of participants varied by quarter and ranged from an estimated 845,980 in the 9th quarter (July through September 2014) to 1,058,405 in the 8th quarter (April through June 2014) (Figure II.A.1). At the end of March 2015, there were an estimated 864,930 indirect program participants for the quarter, which was below the projection of 1,154,011 for the 11th quarter (January through March 2015). (TransforMED's count of indirect program participants is based on the total number of Medicare and Medicaid beneficiaries served at each participating practice.)

Figure II.A.1. Estimated actual enrollment relative to projected enrollment



Source: Review of TransforMED program reports, March 2015.

Note: In program quarter 2, no participants were enrolled. For program quarters 3 and 4, projections are not available.

Note: The awardee used the term *dstimated* enrollment rather than *actual* enrollment because TransforMED enrolled participants indirectly. The awardee used the term *projected* enrollment to indicate its enrollment target .

b. Service measures

TransforMED reached its program process and service delivery goals. Practices participating in the TransforMED program generally focused first on improving patient contact measures (such as number of visits and telephone follow-ups) and process measures (such as number of screenings and number of care plans). A number of these measures were retired when communities reached 90 to 100 percent of their patient panels. Other measures that were targeted and later retired included availability of same-day appointments and availability of extended office hours. TransforMED also retired the incorporation of hospital discharge data into electronic health records (EHRs). As practices implemented processes for using Phytel and Cobalt Talon data, they began to use those data to focus on trying to improve patient outcome measures, such as rates of hemoglobin A1c levels, ED visits, and hospital readmissions.

c. Staffing measures

As of March 2015, TransforMED had not met its goals for hiring staff to provide support to practices and communities with PCMN implementation. TransforMED expected to hire 22.35 full-time equivalent (FTEs) staff for the program from July 1, 2012 to June 30, 2015. As of March 2015, TransforMED hired a cumulative total of 17.35 FTEs since project inception,

meeting approximately 78 percent of its target. Some of the new hires included project managers, facilitators, and one trainer, project data analyst, program director, project control manager, and part-time administrative support staff member. In 2014, 41 super users (facilitators of Cobalt Talon implementation) attended Cobalt Talon training and 35 health coaches attended clinical health coach training; 79 health coaches are participating in online clinical health coach training.

d. Program time line

TransforMED experienced two delays in implementing the program according to its established time line. First, TransforMED launched the program four months later than initially planned due to a waiting period for CMMI to approve the operational plan. Health systems could not fully engage in implementation activities until they received approval of their operational plans. Second, delays in implementing Cobalt Talon and Phytel software occurred due to the unanticipated number and complexity of health information systems requiring integration across participating practices. For some practices, Phytel implementation was delayed up to one year. These implementation delays made it difficult for communities to accomplish as much as they had initially anticipated in terms of developing population management systems and cost management reporting. Despite these challenges, all 77 participating practices completed the implementation process. Thirteen practices did not implement Phytel for various reasons, including having recently implemented an EHR; practices had to have used an EHR for a minimum of one year in order to implement the Phytel software successfully.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external factors. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.4 summarizes the major facilitators and barriers to TransforMED's implementation effectiveness in each domain.

Table II.A.4. Facilitators of and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• Perceived relative advantage	• Design quality
Implementation process		• Self-monitoring/quality improvement • Program resources • Stakeholder engagement
Internal factors	• Implementation climate	• Implementation climate • Internal technological environment
External factors	• General policy environment • External technological environment	• External technological environment • Payment models

Source: Interviews from second site visit, April 2015; document review, March 2015.

a. Program characteristics

One characteristic of the TransforMED program that facilitated PCMN implementation in the communities and participating practices was the perceived relative advantages of aspects of the program over previous practice work processes. These program aspects are the availability of cost and quality data—especially through the implementation of Phytel—and the development of new relationships between entities in the neighborhoods. First, respondents in both communities felt they benefited from the availability of cost and quality data provided by the program and used these data to monitor quality improvements. One community developed a data warehouse and designated physician champions at each participating practice to review metrics and process improvements made across practices each month. This community also developed a monthly scorecard for practices that presented practice- and physician-level performance on a set of quality measures. As one respondent in this community noted during our 2015 site visit, "It [Phytel and Cobalt Talon] prompted us to get familiar with having the claims data and quality data and it forced us to build it into our everyday [operations].… I think that practices have grown to expect these data now. If the [award] went away and we hadn't built the data warehouse, we would be wondering what to do.. ..It just made it part of the standard of things to look at every month. Previously that wasn't a standard." Before PCMN implementation, none of the practices in either community had consistently used patient-level data to track quality indicators such as hemoglobin A1c testing or mammogram screening. As one respondent said during our 2015 site visit, "Phytel was an inspiration for this—knowing your mammogram rate is good information but seeing who is missing and being able to drill down to the patient level is what makes it actionable."

Respondents described Phytel as a robust system that enabled them to organize and report patients' information more efficiently than their previous reporting methods, which is similar to what we heard during our April 2014 site visits. In both communities visited, respondents used Phytel data and reports to either start or improve population management processes. Phytel gave practice staff the ability to review the patients who had care gaps and facilitated outreach to them to close those gaps. For example, one respondent said during our 2015 site visit, "[Before the award,] we were just trying to manage the patients who were sitting in front of us. With the help of Phytel, we could take a bigger look and look at the patients who weren't coming into the office." Also during our 2015 site visit, a respondent at a practice that was already doing population management spoke of how Phytel helped to incorporate population management into its workflow, "We were doing it from an industry standpoint because we knew we should do it, but we didn't truly understand the impact or why ... [we now] see the importance of doing it on a regular basis and keeping track of it.… This project really expanded that to be practice-wide." Respondents from the community that created an internal data warehouse modeled partly off Phytel said that participating in the TransforMED program helped the health system move faster down the path it was already on toward population management, as reflected in a program administrator's comment during our 2015 site visit, "Phytel was really the platform we used to get us to go where we needed to go and see what we needed to build."

Lastly, during our 2015 site visits respondents from both communities spoke of the advantages of developing new relationships with entities in their neighborhoods. For example,

one respondent said, "We are starting to collaborate more with our neighboring hospitals. Before [the award] … they [were] the competition.… [Now] everyone emails each other information.… TransforMED introduced us. They were the glue that brought us together." Another respondent in a different community said that the practice now has contractual relationships with a cardiology group and that the practice is more connected with specialist groups than before the award.

In contrast, in one site we visited, the design of the PCMN program challenged its implementation. TransforMED recommended that many of the PCMH principles be used to support effective implementation of population management systems and cost management reporting. TransforMED reported that it focused on PCMH principles at the start of program implementation to give practices an understanding how of PCMH components are foundational to PCMN, but did not intend to support practices with PCMH certification. However, during our 2015 site visits, some respondents reported that they were told by TransforMED that the program would help them to become certified as PCMHs. As one practitioner said, "If you don't get [PCMH] certified, you don't get any benefits from it. It's a costly endeavor … if you go through all that work, you should be certified." A few respondents noted that, after some time participating in the program, TransforMED said that it would not assist them with PCMH certification, which was frustrating to practices seeking assistance with the certification process. One respondent added that another implementation challenge in the design of the program was that learning how to build and run reports using Cobalt Talon software was too time-intensive for software that would be available for only a three-year period.

b. Implementation process

Three characteristics of the implementation process challenged implementation: the use of data to evaluate where improvements could be made, dedicating resources to program implementation, and engaging stakeholders. Using data from Cobalt Talon to reflect and evaluate where implementation and program improvements could be made was a challenge for the communities. Practices generally relied on the standardized reports developed by Cobalt Talon as opposed to building, customizing, and running their own reports to focus on unique utilization and cost issues in their communities. This was because of the mistrust of Cobalt Talon data that formed when practices saw that claims processing was incomplete even after a three-month lag period, as well as the large investment of time needed to learn how to build and run reports in the software program.

Additionally, participating practices and health systems did not receive funds to support PCMN implementation, and respondents across both communities acknowledged this as a challenge to implementation. Some practices were able to overcome this: for example, one of the communities was part of an accountable care organization (ACO), which committed its own resources to PCMN implementation. The ACO helped hire staff to work with the Phytel reports and to reach out to patients identified as high-risk. Another expense not covered under the award was physicians' time, either to spend more time with patients or to attend PCMN-related meetings. As a practice manager said during our 2015 site visit, "We didn't receive any money. Let's be clear. There was no money. The money makes a big difference. It doesn't have to be a huge amount. I don't think we are asking to pay for a physician's time. Just something that

makes it feel like you can support your physician to attend this meeting. That is important to support this work."

Third, engaging stakeholders was an implementation process factor that challenged PCMN implementation in both communities. PCMN implementation involved changing practice workflows, such as introducing and/or improving preventive care by systematically identifying patients in need of screenings or who would benefit from care plans. Some physicians were hesitant to assign medical assistants greater responsibility for addressing care gaps during patients' visits. During our 2015 site visits, respondents in both communities reported that meeting with physicians and practice staff regularly to discuss workflow changes and demonstrating how these changes led to effective care delivery improved physician and staff engagement over time. For example, in regard to changes in medical assistants' duties, one respondent said that physicians started to respond positively, "[Physicians] understand that it is freeing up their time in the exam room [to spend more time with patients]." Respondents also noted that physicians embraced quality data faster than cost data. They attributed this to physicians working in a clinically driven environment and being more reactive to data that, for example, showed how many of their diabetic patients have high hemoglobin A1c levels, versus cost and utilization information that is generally thought of as being in the realm of the finance department. In addition, physicians expressed concern that there were attribution issues in the new forms of patients' data due to the way the Phytel software defined providers and patient panels and questioned if the data were accurate. Respondents described overcoming resistance to staff engaging in the use of new forms of patients' data by giving them ownership over the resulting improvements, "… it gives everyone some leadership. It makes them feel important. If they feel more informed and more a part of the total office function, each of them understands what each of them does."

c. Internal factors

One internal factor, implementation climate, was a facilitator in one community and created challenges in another. A few respondents in one community described how health system leadership was committed to implementing the PCMN program. The chief executive officer of the health system was committed to strengthening PCMH principles within the community and investing resources in developing care management processes and staff throughout the system, including the non-employed practices participating in the TransforMED program. However, respondents in the other community expressed concern that health system leadership did not take ownership for implementing the program, nor did they initiate collaboration between participating practices in the community to facilitate PCMN implementation.

Another internal factor, the technology infrastructure, challenged program implementation in both communities: the participating practices experienced difficulty implementing Phytel. For the Phytel software to pull data electronically, practices had to ensure that the software mapped to their EHR systems correctly. During both our 2014 and 2015 site visits, some respondents described challenges mapping the software to their EHRs because of how data had been previously entered and where the data were located. For example, respondents at one practice said that practice staff spent a significant amount of time creating data flow sheets to link their EHR to Phytel and that, as a result, they were left with little time to use Phytel to develop

population management workflows. One of these respondents said, "A big part of our time after we got Phytel was re-recording our data ... that also means we were taking a long time to get to the starting line. We have not gotten to use Phytel in a way we would have liked to." In both communities, a few respondents at practices that did not correctly map the software to their EHRs used inaccurate Phytel data on care gaps to notify patients they believed were due for services but in reality were not. As a result, a number of patients expressed frustration with the practices.

d. External factors

Features of the environmental context in which the organization is located can also influence program implementation. One external factor, the general policy environment, facilitated PCMN implementation in both communities. During our 2015 site visits, respondents in both communities spoke of how the goals of the PCMN program aligned with goals they were working toward through other initiatives. Respondents in one of the communities described how their participation in an ACO facilitated PCMN implementation because of the alignment between the quality measures they tracked for the PCMN program and those they tracked for the ACO. In the other community, one respondent spoke of how working with Phytel and establishing population management systems aligned with the goals of two health insurance company initiatives in which the practice participated.

Another external factor, the technology environment, facilitated PCMN implementation in one community and challenged it in the other. In one community, the health system and participating practices utilized the state's health information exchange (HIE) to review data on patients who visited the ED or were admitted and discharged from the hospital. Respondents, during our 2015 site visit, described how before the HIE was established, their health system had relied on claims data with a two- to six-month lag to identify patients admitted to the ED or hospital. One respondent said, "[You would say to the patient,] 'You went to the ED a lot three months ago!' By then you [the health system] missed your chance to make a change because it's [the medical situation] is resolved by that time. You need to do something during your [the patient's] first visit [to the ED or hospital], not the fifth." However, in the other community the health system and participating practices did not utilize the state's HIE to obtain or review data because the state's technology environment was weak: "The state doesn't have a ton of money and a big organized plan. They still have broadband in the rural areas. That robust health IT infrastructure isn't there." In addition, the respondent explained that there was no incentive for the health system to facilitate connection among providers in the state without a robust state HIE: "Is the technology, both from a cost and utility perspective, coming…to a point where we can start to…get more wired to these other providers...[and] make that investment? The answer is not yet. You will not see … [the health system] spend millions of dollars to wire all these places without a robust state HIE.… When you participate in a project like this [the TransforMED program], you start to think about the degree to which we, as a [health] system, should take it upon ourselves versus funds flowing in from the outside.… From a health information exchange perspective, we could all start connecting [but] where is the state in that?"

Another external factor, payment models, challenged TransforMED program implementation in one community. Two respondents spoke of lack of reimbursement for

adopting PCMH concepts during our 2015 site visit. For example, one respondent noted that, after adopting PCMH concepts, the average number of patients the practice sees in a day decreased because PCMH-related activities increased physicians' workloads: "It cost us a third of the production throughput in order to do PCMH.… It's very expensive to do … [and] no one else is paying a thing for it." One respondent noted that the introduction of transition payment codes were helpful, but that better reimbursement from payers is still needed.

4. Sustainability and scalability

TransforMED included in its original PCMN program design plans for spreading the PCMN program to 18 to 22 additional practices within each community. The focus of the spread was on care management for high- and rising-risk patients, access, care coordination across the medical neighborhood, and continuous performance improvement. The spread did not include the use of the Phytel or Cobalt Talon software systems because of the costs associated with implementation and training. Various communities used different strategies to spread the PCMN program to additional primary care practices. Many communities incorporated the additional practices in the PCMN program during the initial stages of implementation (Note: the spread practices were not included in the count of the 90 participating practices.) Other communities identified practice coaches to work with the additional practices to spread the PCMN program. However, committing resources to practice coaches to work with practices not affiliated with the health system proved to be challenging. According to TransforMED, gleaning information from their interactions with practices, they saw the cross-community collaboration that began in Year 3 of the award as being the most successful mechanism for sustaining the PCMN program within communities and spreading it to other practices.

There were a number of payment models across the PCMN communities. Within the fee-for-service environment, TransforMED tried to help practices identify value-based revenues, such as the Medicare chronic care management fees and wellness visits. TransforMED also helped practices identify where closing care gaps could result in service reimbursements. For practices that were part of a hospital system, TransforMED worked with practices and payers to negotiate reduced hospital readmission penalties.

Respondents from both communities at which the 2014 and 2015 site visits were conducted were committed to sustaining Phytel or a comparable population management system, but described having limited resources to sustain Phytel or Cobalt Talon. To offset this, one of the health systems developed its own version of Phytel and Cobalt Talon. The other health system was negotiating the cost of maintaining the software system with Phytel. Respondents in both communities commented about the difficulty they would have if they did not have the capabilities that Phytel provided: "Once a clinician knows that 250 patients are not getting the care they should get, you have a duty to fix that problem. I see us staying committed to using a product like Phytel to scale up." Neither community planned to continue using the Cobalt Talon cost management system. For one community, this was because the community already had the same data that Cobalt Talon provided in its financial system. For the other, it was because their internal data warehouse served as a replacement for both Cobalt Talon and Phytel.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected health system and clinical sites provided important insights into the implementation process. Although these in-person interviews provide a rich source of data, views from the leadership and staff are limited to a small number of clinical locations and might not reflect the perspectives of clinicians practicing at other sites. To assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in fall 2014, the third year of the HCIA-funded program. Data from the survey provide additional insights into the implementation process and experience as well as the contextual factors that might affect implementation effectiveness at TransforMED.

In this section, we report on the views of daily work life and practice of clinicians at practices that are participating in the TransforMED program (hereafter referred to as TransforMED clinicians). First, we focus on the contextual factors that can affect program implementation, including the characteristics of the practice locations, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care team functions. We then present data on the alignment of the TransforMED clinicians' views and experiences with the overall goals of the HCIA-funded innovation, as well as their awareness of and participation in the program and their views of the barriers to and facilitators of successful program implementation. Note that throughout this section, the number of clinicians in each response category does not always sum to the total number of TransforMED respondents due to survey item nonresponse, as well as clinicians who reported that a given question did not apply to their practice and thus did not provide a response.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

A total of 323 TransforMED clinicians responded to the survey (resulting in a response rate of 70 percent). Of the respondents, 234 were physicians, 54 were nurse practitioners, and 28 were physician assistants. These clinicians practiced predominantly at clinical locations with three or more clinicians (67 percent). Other clinical locations included the following: Federally Qualified Health Centers (FQHCs) (5 percent) or other community health centers (10 percent), a hospital run by a private for-profit or not-for-profit organization (6 percent), a medical school or university (2 percent), a group or staff model health maintenance organization (1 percent), a solo practice (1 percent), and a two-clinician practice (4 percent). Most TransforMED clinicians reported that their primary source of compensation was a salary adjusted for performance (57 percent).

TransforMED clinicians reported working in settings that are advanced in terms of health IT. This aligns with the participation requirements for the TransforMED program; clinical sites were required to have used an EHR for at least one year before program implementation to support the implementation of the population management software from Phytel. Nearly all clinicians reported using electronic systems for ordering tests and procedures (97 percent),

accessing laboratory test results (98 percent), prescribing medications (94 percent), drug dosing and interaction alerts (98 percent), or entering clinical notes (97 percent). In addition, clinicians reported using electronic referral tracking systems (84 percent) and patient registries (a function of the Phytel population management software) (80 percent), functions that are generally advanced and not in widespread use nationally (DesRoches, Painter, and Jha 2014). TransforMED clinicians also reported that they offer patient-facing technologies. About 68 percent of clinicians offered their patients the option to request an appointment online, 76 percent offered patients the ability to request prescription refills online, and 72 percent offered them the ability to email a clinician about a medical question or concern.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall careers, level of burnout, and perceptions of their practice environments can all have an effect on the success of program implementation and organizational change. As shown in Table II.B.1, TransforMED clinicians are generally satisfied with their careers in medicine. However, only 34 percent reported being very satisfied and more than one-fourth were experiencing some symptoms of burnout at the time the survey was taken.

Table II.B.1. Career satisfaction and burnout

Survey item	Number of respondents	Percentage of respondents
Overall satisfaction with career		
Very satisfied	111	34%
Somewhat satisfied	154	48%
Neither	14	4%
Somewhat dissatisfied	33	10%
Very dissatisfied	--	--
Degree of burnout		
I enjoy my work. I have no symptoms of burnout.	44	14%
Occasionally I am under stress, and I don't always have as much energy as I once did, but I don't feel burned out.	167	52%
I am definitely burning out and have one or more symptoms of burnout, such as physical and emotional exhaustion.	79	24%
The symptoms of burnout that I'm experiencing won't go away. I think about frustrations at work a lot.	26	8%
I feel completely burned out and often wonder if I can go on. I am at the point where I may need some changes or may need to seek some sort of help.	--	--

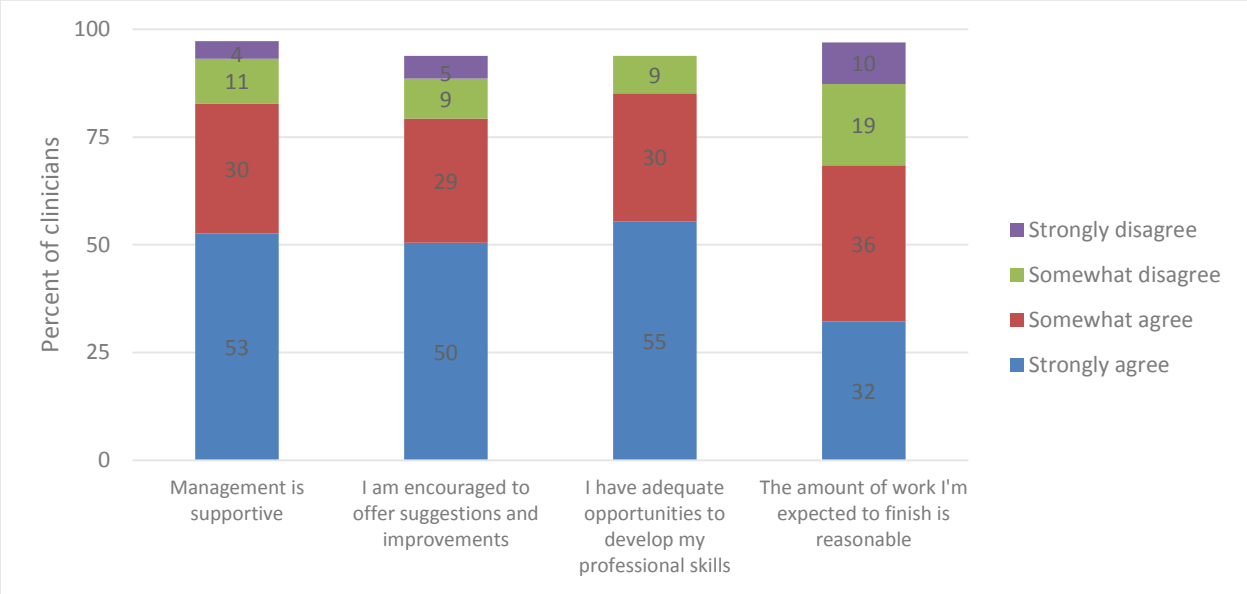
Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Column totals might not sum to 100 percent due to rounding. Survey items with fewer than11 respondents are not shown because of confidentiality restrictions.

The clinicians gave similar ratings to their workplace management. As shown in Figure II.B.1, half or more of responding clinicians strongly agree that their management team was supportive, that they were encouraged to offer suggestions and improvements, and that they had adequate opportunities for professional development; about a third strongly agreed that the amount of work they were expected to complete each day was reasonable.

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. For example, 25 percent of responding clinicians strongly agreed and 39 percent somewhat agreed with the statement "It is possible to provide high quality care to all of my patients." Among the major barriers to providing optimal care reported by most TransforMED clinicians were insufficient reimbursement, lack of time to spend with patients during visits, and lack of timely information about care provided to patients by other physicians.

Figure II.B.1. Workplace ratings



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Totals might not sum to 100 percent due to rounding and item non-response. Survey items with fewer than11 respondents are not shown because of confidentiality restrictions.

c. Clinicians' perceptions of care team functioning

A large majority (88 percent) of TransforMED clinicians reported working as part of a care team and, overall, their perceptions of how these teams function was positive. TransforMED clinicians mostly agreed that members of the care team relayed information in a timely manner (97 percent), had sufficient time for patients to ask questions (89 percent), used common terminology when communicating with each other (95 percent), verbally verified information they received from each other (84 percent), and followed a standardized method of sharing information when handing off patients (78 percent).

d. Alignment with goals of primary care redesign

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The inclusion of the extremely important category helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are merely important. In Table II.B.2, we present results based on the proportion of clinicians rating each of these goals as extremely important. The views of TransforMED clinicians generally aligned with the goals of PCR. Most clinicians rated 8 of the 13 goals as extremely important. It is not surprising that about one-third of clinicians ranked increased use of EHRs and other health IT as important, given that all of 90 practices were already using EHRs and were required to have done so for a minimum of one year before Phytel software implementation.

Table II.B.2. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Reducing ED visits	222	69%
Reducing hospital readmissions	204	63%
Improving patients' capacity to manage their own care	201	62%
Increasing access to primary care	200	62%
Improving care coordination for patients with chronic conditions	191	59%
Improving care continuity in primary care	188	58%
Reducing overall health care spending	184	57%
Improving appropriateness of care	180	56%
Increasing the use of evidence-based practices in clinical care	157	49%
Improving capability of health care organizations to provide team based care	131	41%
Improving the capability of health care organizations to provide patient-centered care	129	40%
Increasing use of EHRs and other health IT	102	32%
Increasing the number of primary care practices functioning as a PCMH	101	31%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Percentages are calculated as the percentage of total respondents who rated each item as "extremely important." Items are rated separately; percentages do not add up to 100 percent.

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the TransforMED program is to change the way patients' information is organized and used in primary care delivery. Understanding clinicians' perceptions of using cost and quality information could be helpful in understanding the effect of the program on patients' outcomes. For example, the lead clinician at one clinical site believed that population management software was necessary to quickly and easily identify gaps in patients' care, and that it facilitated outreach to patients to raise their awareness of these care gaps and to remind them to visit the clinic. However, practice staff at participating sites primarily used the software

and most clinicians did not directly interact with it. Those clinicians who felt more negatively about using cost and quality information might be less likely to enthusiastically implement the intervention. In this section, we report on TransforMED clinicians' experiences with and perceptions of the program.

a. Awareness of the program and receipt of training

Only 57 percent of the TransforMED clinicians we surveyed were at least somewhat familiar with the program. Although surprising, this might indicate that the program was implemented largely by practice management and allied health professionals, such as care managers, so that practitioners were unaware of a formal program, or because of resistance from clinicians to change workflows or use new tools. It might also be because some practices are participating in other initiatives that are conducting similar activities, and clinicians do not have in-depth familiarity with the programs to be able to differentiate between them.

b. Perceived effect of program on patients' care

Clinician's perceptions of the effect of the program on the care they provide to patients were mixed. More than half of clinicians who were familiar with the program believed it would have a positive effect on the patient-centeredness of the care they provide (56 percent). Fewer than half of physicians familiar with the TransforMED program believed the program would have a positive effect on the ability to respond to patients in a timely manner (34 percent) and the quality (47 percent), efficiency (31 percent), safety (41 percent), and equity of care (26 percent). Few clinicians perceived an actual negative impact of the program; rather, they believed the intervention would have no effect on the care they provide or that it was simply too soon to tell.

c. Barriers to and facilitators of program implementation

Finally, we asked TransforMED clinicians who were at least somewhat familiar with the program to rate the effect of a series of barriers and facilitators to program implementation. Clinicians were asked about the perceived effect of the TransforMED program and the barriers to and facilitators of implementation only if they reported being at least somewhat familiar with the program. The most often-cited facilitator to program implementation was the quality of interpersonal communications with other allied health professionals; however, fewer than half of the clinicians who were familiar with the program said this was a facilitator (Table II.B.3.). The most often-cited barrier to program implementation was the amount of required documentation.

Table II.B.3. Barriers and facilitators to program implementation

	Positive impact		No impact		Negative impact		Not applicable/don't know	
Survey item	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Level of program funding	51	28%	34	18%	12	6%	84	45%
Amount of required documentation	26	14%	52	28%	74	40%	31	17%
Availability of personnel	52	28%	61	33%	34	18%	36	19%
The amount of time required by the program	29	16%	78	42%	43	23%	33	18%
Availability of relevant patient information at the point of care	77	42%	53	29%	17	9%	36	19%
Required use of computer and communications technology	77	42%	44	24%	33	18%	27	15%
Availability of evidence-based clinical information	71	38%	66	36%	--	--	33	18%
Availability of community resources to care for patients with complex conditions	66	36%	62	34%	15	8%	40	22%
Quality of interpersonal communications with other providers	75	41%	59	32%	--	--	40	22%
Quality of interpersonal communications with specialists	62	34%	66	36%	--	--	48	26%
Quality of interpersonal communications with other allied health professionals	83	45%	44	24%	--	--	46	25%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Row totals might not sum to 100 percent because of rounding. Figures are based on the number of clinicians who reported being at least somewhat familiar with the TransforMed program. Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions.

4. Conclusions about clinicians' attitudes and behavior

TransforMED assisted participating health systems and primary care practices with the implementation of population management systems and adoption of cost-management reporting, in addition to facilitating collaboration with other participating practices through shared learning activities. Participating practices used a variety of implementation strategies to implement the program. Almost half of all participating clinicians said they were unfamiliar with the TransforMED program. However, despite their lack of familiarity with the program, most clinicians rated most of the goals of the primary care redesign intervention as extremely important.

C. Impacts on patients' outcomes

1. Introduction

In this part of the report, we draw preliminary conclusions based on available evidence about the impacts of TransforMED's HCIA program on patients' service use. Although TransforMED's program serves Medicaid beneficiaries and Medicare beneficiaries enrolled in managed care plans as well as Medicare FFS beneficiaries, due to limiatations in available data we have analyzed outcomes only for the Medicare FFS population (including those who are dually eligible for Medicare and Medicaid). Results might not be generalizable to the full population that TransforMED's program serves.

In this section, we first describe the methods for estimating impacts (Section II.C.2) and then the characteristics at the start of the intervention of the 87 practices (of 90) that we include in the treatment group for the impact analysis (Section II.C.3). We next demonstrate that the 87 treatment practices were similar at the start of the intervention to the 290 practices we selected as a comparison group, which is essential for limiting potential bias in impact estimates (Section II.C.4). Finally, in Section II.C.5, we describe the quantitative impact estimates, their plausibility given implementation findings, and our conclusions about program impacts on service use. Our conclusions in this report are preliminary because the analyses do not cover the full time period that we will include in the final impact analysis in future reports. In future reports, when data for the full time period become available, we will also estimate impacts on spending in addition to service use.

2. Methods

a. Overview

We estimated program impacts as the difference in outcomes for patients assigned to the 87 treatment practices and outcomes for patients assigned to 290 matched comparison practices, adjusting for any differences between the groups before TransforMED's HCIA intervention began. To focus the analyses, we specified a limited number of primary tests before examining any impact results and gave the awardee and CMMI an opportunity to review them. Each primary test defined an outcome, population, time period, and direction of expected effects for which we hypothesize impacts if the program is effective. We drew conclusions about impacts based on the results of these primary tests and the plausibility of the primary test results with the implementation findings and secondary quantitative tests (robustness and model checks).

b. Treatment group definition

We defined the treatment group separately in each of four baseline quarters before the program began on January 1, 2013 (the baseline period), and in each of 8 intervention quarters after the program began (the intervention period). We were able to include only 87 of the 90 participating practices in the treatment group. Three participating practices were dropped from analysis because we were unable to attribute Medicare beneficiaries for several program quarters. In each quarter of the baseline or intervention period, the treatment group consisted of Medicare FFS beneficiaries who (1) were attributed (see below) to one of the 87 treatment practices on or before the first day of the quarter and (2) had observable outcomes for at least one day in the quarter. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

Using this definition of the treatment group, a beneficiary who has previously been assigned to the treatment group will *remain* a member of the treatment group for the rest of the relevant period (baseline or intervention), as long as he or she is still enrolled in Medicare FFS. This definition ensures that, during the intervention period, beneficiaries do not exit the treatment group solely because the intervention succeeded in reducing their service use (including visits at treatment practices). The definition for the baseline period corresponds to that of the intervention period so that, across the two periods, interpretation of the population changes over time should be comparable.

Attribution. We attributed beneficiaries to practices using an algorithm similar to that used by CMMI for the Comprehensive Primary Care (CPC) Initiative. Specifically, in each baseline and intervention month, we attributed beneficiaries to the primary care practice whose providers (physicians, nurse practitioners, or physician assistants) provided the plurality of primary care services in the past 24 months. When there were ties, we attributed the beneficiary to the practice he or she visited most recently. This attribution method required identifiers for the practice site or the providers who worked in the treatment practices (and when), as well as identifiers for providers in other practices in the comparison regions who could compete for beneficiaries (when determining which practice provided the plurality of primary care services). TransforMED provided identifiers for the treatment providers; we purchased data on providers in nontreatment practices from SK&A, an outside health care data vendor that maintains and verifies lists of providers who work in practices throughout the country. For FQHCs and Rural Health Centers (RHCs), we obtained *CMS Certification Numbers* from the Integrated Data Repository for all such health centers in the five states in which FQHCs and RHCs participated in the TransforMED program.

Definition of high-risk subgroup. Because some aspects of TransforMED's intervention (including care management) focuses on improving care for beneficiaries at rising or high risk of hospitalization and other expensive care, we also defined a high-risk subgroup of the treatment group each quarter. For each baseline quarter, this subgroup consists of the beneficiaries with a January 2012 Hierarchical Condition Category (HCC) score in the top quarter among all beneficiaries ever seen by a treatment practice during the baseline period, by market area. In each intervention quarter, the high-risk subgroup consists of beneficiaries whose HCC score was in the top quarter, by market area, among all observable treatment group members at the start of

the intervention period. The HCC score is a continuous variable that predicts a beneficiary's Medicare spending in the following year relative to the national average, with 1.0 indicating that the predicted spending is at the national average and 2.0 indicating that it is twice that average.

c. Comparison group definition

The comparison group consists of Medicare FFS beneficiaries assigned to 290 matched comparison practices during each quarter of the baseline and intervention periods. We selected comparison practices that were similar during the baseline period to the treatment practices in factors that can influence patient outcomes, especially those that TransforMED used when deciding which practices to recruit for the intervention. This section describes how we constructed the matched comparison group whereas Section II.C.4 shows the balance we achieved between the two groups on the matching variables.

We selected the 290 comparison practices in four steps. First, we identified market areas from which to draw potential comparison practices. We chose the entire state for 3 of the 15 TransforMED program sites (Kansas, Mississippi, and Nebraska). For each of the remaining 12 sites, we selected a within-state region or, for one site that had treatment practices in two states (Kentucky and Indiana), a region that includes a portion of both states. In all cases, we balanced the need for a large pool of comparison practices to ensure a sufficient sample of well-matched comparison practices against the desire to restrict the pool to potential comparison practices located in areas similar to those of treatment practices, ensuring face validity of our approach.

Second, we constructed matching variables, defined at the start of the intervention period (January 1, 2013), for all treatment and potential comparison practices. These variables included characteristics of all Medicare FFS beneficiaries assigned to the practices (for example, mean HCC score and utilization in the baseline period); characteristics of high-risk beneficiaries assigned to the practices; characteristics of the geographic location of the practices; and, for nonhealth centers, characteristics of the practices overall (for example, the number of providers in the practice or whether the practice is owned by a hospital or health system). Section II.C.4 shows the matching variables and their data sources.

Third, we narrowed the pool to 7,380 potential comparison practices by excluding those practices that were (1) participating in one of the three other federal primary care initiatives that were operating in the 15 TransforMED market areas (the Multi-Payer Advanced Primary Care Practice [MAPCP] Demonstration, the Comprehensive Primary Care [CPC] Initiative, and the Federal Qualified Health Center [FQHC] Demonstration);[1] (2) owned by one of the 15 participating health systems; (3) recruited by TransforMED during the second phase of the HCIA program as part of its efforts to expand the PCMN program's reach by 18 to 22 additional practices within each community; (4) had an average of fewer than 25 assigned Medicare FFS beneficiaries during the four baseline quarters; and (5) had a practice size of 100 or more total providers.

[1] The exception to this is in Michigan, where three of the five treatment practices are participating in MAPCP. We did not exclude practices participating in the MAPCP initiative from the potential comparison pool in Michigan.

Finally, we used propensity scores to select 290 comparison practices (from the pool of 7,380) that were similar to the 87 treatment practices on the matching variables. The propensity score for a given practice is the predicted probability, based on all matching variables, that the practice is part of the treatment group (Stuart 2010). The score collapses information from all of the matching variables into a single number for each practice that we used to assess how similar practices are to one another. We matched each treatment practice to one or more comparison practices with similar propensity scores, with the aim of generating a comparison group that is similar, on average, to the treatment group on the matching variables (see Section II.C.4 to assess balance between treatment and comparison groups after matching). The propensity-score matching approach, however, does not ensure that each comparison practice matches exactly to its treatment practice on all matching variables.

We ran two separate propensity-score matching models—matching health centers separately from nonhealth centers because the variables available for matching these two groups differed slightly. Within each propensity-score model for matching, we further required that a treatment practice could match only to a comparison practice located in the same market area. We required each treatment practice to match to at least one, but no more than five, comparison practices and that the ratio of comparison to treatment practices be at least 3:1.

After completing the matching, we reviewed the list of selected comparison practices and removed any that seemed qualitatively unlike the target practices for the HCIA intervention—that is, Indian Health Services and walk-in clinics. We then assigned Medicare FFS beneficiaries to the comparison practices in each intervention quarter using the same rules we used for the intervention group (see Section II.C.2.b). Further, we defined a high-risk subgroup of comparison members in each quarter using the same rules as for the treatment group (that is, using the distribution of HCC scores observed among beneficiaries seen in a treatment practice during the start of the intervention period).

d. Construction of outcomes and covariates

We used Medicare claims from January 1, 2009, to December 31, 2014, for beneficiaries assigned to the treatment and comparison practices to develop two types of variables: (1) **outcomes**, defined for each person in each baseline or intervention quarter that the person was a member of the treatment or comparison group, and (2) **covariates** that describe a beneficiary's characteristics at the start of the baseline and intervention periods, and are used in the regression models for estimating impacts to adjust for beneficiaries' characteristics before the period began. We used one set of baseline covariates, without updating them each quarter, to avoid the potential bias that could occur if the intervention affected both control variables and outcomes. For example, the intervention could result in greater contact with the health system and earlier diagnoses of diseases and conditions, which could affect both health-related characteristics and outcomes. If we adjust for changes in health-related status during the intervention period, we can adjust away part of the impact of the intervention. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. We calculated two quarter-specific outcomes to measure service use:

1. All-cause inpatient admissions (number/quarter)
2. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission

Both outcomes are outcomes that CMMI has specified as "core" for the evaluations of all HCIA programs. Two additional core outcomes—the number of unplanned inpatient readmissions within 30 days and total Medicare Part A and B spending—are assessed in our quarterly reporting to CMMI, but are not included here. We do not include spending in our primary tests this year (preliminary findings) because the awardee expected its largest impacts on spending only after its HCIA program had operated for two years; we will assess impacts on spending in our next annual report, when data are available for the final quarters of the TransforMED intervention, which ended June 30, 2015. We do not include readmissions in our primary tests this year because the awardee did not explicitly expect to affect this outcome.

Covariates. The covariates, or predictor variables, include (1) whether a beneficiary has each of 18 chronic conditions (including physical health, mental health, and disabilities), created by applying Chronic Condition Warehouse algorithms to claims in the 12 to 36 months (depending on the condition) before the start of the baseline or intervention periods; (2) HCC scores; (3) demographics (age, gender, and race or ethnicity); (4) whether the beneficiary is dually eligible for Medicare and Medicaid; (5) whether the beneficiary is a member of the high-risk subgroup, and (6) original reason for Medicare entitlement (old age, disability, or end-stage renal disease).

e. Regression model

We used a regression model to implement the difference-in-differences design for estimating impacts. For each quarter-specific outcome, the model estimates the relationship between the outcome and a series of predictor variables, assuming that each one of the predictor variables has a linear (additive) relationship with the outcome. The predictor variables include the beneficiary-level covariates (defined in Section II.C.2.d); whether the beneficiary is assigned to a treatment or a comparison practice; an indicator for each practice (which accounts for stable differences among practices in their outcomes over time); and an interaction of a beneficiary's treatment status with each post-intervention quarter. The estimated relationship between the interaction term and outcomes in a given quarter is the impact estimate for that quarter. It measures the average difference between outcomes for beneficiaries assigned to the treatment and comparison practices that quarter, subtracting out any differences between these groups during the four baseline quarters. By providing separate impact estimates for each intervention quarter, the model allows the program's impacts to change the longer the practices are enrolled in the program (which is expected to occur). We can also test impacts over discrete sets of quarters, which is needed to implement the primary tests discussed in the next section. Finally, the model quantifies the uncertainty in the impact estimates, facilitating statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. The model used robust standard errors to account for clustering of

outcomes across quarters for the same beneficiary and a dummy variable for each practice (fixed effects) to implicitly account for clustering of outcomes for beneficiaries assigned to the same practice. Appendix 2 provides details on the regression methods, including descriptions of the weights each beneficiary receives in the model.

f. Primary tests

Table II.C.1 shows the primary tests for TransforMED, by domain. We have specified our primary tests based on the data we expect to be available for the third annual report (due August 2016) and for all outcomes to be considered, whether examined in this or future reports. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important (expressed as a percentage of change from the counterfactual—that is, the outcomes that beneficiaries in the treatment group would have had if they had not received the treatment). The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for detail and a description of how we selected each test).

Our rationale for selecting the primary tests in Table II.C.1 is as follows:

- **Outcomes.** TransforMED's central goal was to decrease total Medicare and Medicaid spending by 4 percent by Year 3 of the program. For this reason, we chose to analyze impacts on Medicare Part A and B spending. Reductions in hospitalizations and ED visits are identified as primary drivers that will enable these cost reductions. Therefore, we selected primary tests examining hospitalizations, ED visits, and Medicare Part A and B spending.
- **Time period.** TransforMED expects to have measurable impacts on spending by the third year of the program (few impacts are expected in the first two years of practice participation). Given this, we chose to analyze impacts for spending during the final two quarters of the program's operation (that is, intervention quarters 9 through 10), as these correspond to the third year of the program, following two complete years of program operation. We chose to analyze impacts for hospitalization and ED visits over a longer period—namely, the final year of the practice participation (that is, intervention quarters 7 through 10)—because reductions in these outcomes are expected to occur earlier, as practices began using the cost- and population-management data to better manage their patients' care. In this report, because we have data only through December 2014 (intervention quarter 8), we present preliminary results for hospitalizations and ED visits over intervention quarters 7 and 8 only (not intervention quarters 9 and 10), and we do not present any results for spending (which we plan to assess in the future over intervention quarters 9 and 10).

Table II.C.1. Specification of the primary tests for TransforMED

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population	Substantive threshold (impact as percentage of the counterfactual)[c,d]
Quality-of-care outcomes (0)	n.a.—Awardee does not explicitly plan to affect quality-of-care outcomes	n.a.	n.a.	n.a.
Service use (4)	All-cause inpatient admissions (#/person/quarter)	Average over intervention quarters 7 through 10	All Medicare FFS beneficiaries attributed to treatment practices	-5%
	Outpatient ED visit rate (#/person/quarter)			-5%
	All-cause inpatient admissions (#/person/quarter)		Medicare FFS high-risk beneficiaries attributed to treatment practices	-15%
	Outpatient ED visit rate (#/person/quarter)			-15%
Spending (2)	Medicare Part A and B spending ($/person/month)	Average over intervention quarters 9 through 10	All Medicare FFS beneficiaries attributed to treatment practices	-3%
	Medicare Part A and B spending ($/person/month)		Medicare FFS high-risk beneficiaries attributed to treatment practices	-15%

Notes: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

High-risk beneficiaries are defined as beneficiaries with a Hierarchical Condition Category score in the top quarter among all beneficiaries seen by treatment practices during the period (baseline or intervention), by market area.

[a] We have adjusted primary test results for the multiple comparisons made within the service use domain. In future reports, we will adjust within each domain (service use or spending), but not across domains.

[b] The regression models control for differences between the treatment and comparison groups during the baseline year when estimating program impacts.

[c] For total Medicare FFS spending among the full population, we set the substantive threshold to 75 percent of TransforMED's expected effect (see Section II.C.2.f). For the other outcomes or subpopulations, for which TransforMED did not set an explicit target, we set the threshold equal to reductions in acute care use or spending that Peikes et al. (2011) indicated could be feasible among either high-risk or general-population beneficiaries (as applicable) in a patient-centered medical home program.

[d] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention.

ED = emergency department; FFS = fee-for-service.

n.a. = not applicable.

- **Population.** The most inclusive definition possible would be to use all Medicare FFS beneficiaries attributed to the treatment practices, and to compare them with all Medicare FFS beneficiaries attributed to the comparison practices; the intervention is designed to affect the care of patients of all risk levels. However, although the program does not explicitly target a specific population for services, TransforMED's impacts on the outcomes of interest should be concentrated among high-risk beneficiaries, both because there are more opportunities to reduce acute care for this high-risk population and because beneficiaries in this group are more likely to receive intensive interventions, such as case management. Because there are trade-offs between analyzing a high-risk subpopulation (for which expected effects would be larger but the sample size is moderate) and analyzing the entire Medicare FFS population (which is more representative of the program population served but with smaller anticipated effects), we chose to assess both in our primary tests. Our definition for high-risk beneficiaries is described in Section II.C.2.b.
- **Direction (sign) of the impact estimate.** The primary test is to check for a reduction relative to the counterfactual of each of the outcome measures.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we have specified thresholds for what we call substantive importance. For the full population, the 3 percent threshold we chose for substantive importance on spending is 75 percent of TransforMED's expected effect on this outcome in the third year of practice participation. (We use 75 percent recognizing that TransforMED could still be considered successful if it came close to, but did not achieve, its fully anticipated effects.) The awardee did not specify anticipated impacts on the intermediate outcomes of hospitalizations and ED visits or among subpopulations, so all of our other thresholds—for spending among a high-risk population, outpatient ED visits among either a high-risk or full population, and all-cause admissions among either a high-risk or full population—were instead taken from the literature (Peikes et al. 2011). These thresholds were based on the assumption that a successful primary care intervention could cause a reduction in spending or service use of 5 percent among a general population and 15 percent among a high-risk population.

Due to limitations in data availability, we were able to conduct the primary tests in this report only partially. Specifically, because we have data for only eight intervention quarters, we estimated impacts only for hospitalizations and ED visits and included only intervention quarters 7 through 8. The third annual report will include spending and all intervention quarters of the primary tests (7 through 10 for ED visits and hospitalizations; 9 and 10 for spending).

g. Secondary tests

We also conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the non-experimental design of our study or random fluctuations in the data. We will have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. Specifically,

we estimated the program's impacts on hospitalizations for both the full population and the high-risk beneficiaries during four additional intervention quarters—that is, the first 12 months of program operation (intervention quarters 1 through 4). Because we and TransforMED expect program impacts to increase over time, the following pattern would be highly consistent with an effective program—largest impacts in quarters 7 and 8 (which is the time period for the primary tests in this report), and smaller impacts during quarters 1 through 4. In contrast, if we found larger differences in outcomes (favorable or unfavorable) in the first year of the program than in quarters 7 and 8, this could suggest a limitation in the comparison group, not true program impacts.

h. Synthesizing evidence to draw conclusions

Our conclusions about program effectiveness are based on the primary test results, the results of secondary tests, and the plausibility of those findings given the implementation evidence. The four possible conclusions we allowed ourselves are as follows: (1) statistically significant favorable effect (the highest level of evidence), (2) substantively important favorable effect, (3) indeterminate effect, and (4) substantively important unfavorable effect. (We cannot conclude that a program has a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which do not test for evidence of program harms.)

Our decision rules for each of the four possible conclusions are described in Appendix 3. In short, we concluded that a program had a statistically significant favorable effect if (1) at least one primary test result was favorable and statistically significant, after adjusting the statistical tests to account for multiple tests (if applicable) within the domain; or (2) the average impact estimate across all primary tests in the domain was favorable and statistically significant. In both cases, we also had to determine that the primary test results were plausible given the secondary tests and implementation evidence. We concluded that a program had a substantively important favorable effect if the average impact estimate was substantively important but not statistically significant, and if the result was plausible given the secondary tests and implementation evidence. In contrast, if the average impact estimate was unfavorable (opposite the hypothesized direction), larger than the substantive threshold, and unfavorable effects were plausible given the other evidence, we concluded the program had a substantively important unfavorable effect. Finally, if the tests in a domain did not meet any of these criteria, we concluded that the impact was indeterminate.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the treatment group at the start of the intervention (January 1, 2013), which can be seen in the second column of Table II.C.2. (Table II.C.2 also serves a second purpose—to show the equivalence of the treatment and comparison practices at the start of the intervention—which we describe in Section II.C.4.). For benchmarking purposes, the last column shows the values of relevant variables for the national Medicare FFS population, when available.

Table II.C.2. Characteristics of treatment and comparison practices when the intervention began (January 1, 2013)

Characteristic of practice	Treatment practices (N = 87)	Unmatched comparison pool (N = 7,380)	Matched compar-ison group (N = 290)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Exact match variables[c]						
Health center (%)	11.5	5.3	11.5	0.00	0.00	n.a.
Market area (%)						
Alabama	6.9	8.1	6.9	0.00	0.00	n.a.
Connecticut	6.9	8.1	6.9	0.00	0.00	n.a.
Florida	5.8	15.8	5.8	0.00	0.00	n.a.
Georgia	9.2	7.1	9.2	0.00	0.00	n.a.
Indiana	9.2	8.3	9.2	0.00	0.00	n.a.
Kansas	6.9	4.7	6.9	0.00	0.00	n.a.
Kentucky/Indiana	4.6	1.6	4.6	0.00	0.00	n.a.
Maryland	8.1	10.7	8.1	0.00	0.00	n.a.
Massachusetts	8.1	7.2	8.1	0.00	0.00	n.a.
Michigan	2.3	2.3	2.3	0.00	0.00	n.a.
Mississippi	8.1	6.1	8.1	0.00	0.00	n.a.
Nebraska	6.9	5.1	6.9	0.00	0.00	n.a.
North Carolina	6.9	7.0	6.9	0.00	0.00	n.a.
Oklahoma	5.8	5.7	5.8	0.00	0.00	n.a.
West Virginia	4.6	3.2	4.6	0.00	0.00	n.a.
Propensity-matched variables[d]						
Characteristics of a practices location(s)						
Located in an urban zip code (%)	79.3	83.9	76.3	3.01	0.07	NA
Medicare Advantage penetration rate (2011) (%)	17.5	19.4	17.8	-0.34	-0.04	NA
Located in a Health Professionals Shortage area (primary care) (2011) (%)	11.5	20.1	15.7	-4.16	-0.12	NA
Characteristics of all patients attributed to practices during the baseline year (January 1, 2012 – December 31, 2012)						
Number of beneficiaries	970	406	1048	-77.80	-0.10	n.a.
HCC risk score	1.12	1.18	1.11	0.01	0.03	1.0
All-cause inpatient admissions (#/1,000 patients/quarter)	79.45	90.35	80.24	-0.79	-0.04	74[e]
Outpatient ED visit rate (#/1,000 patients/quarter)	138.22	142.09	133.32	4.90	0.09	105[f]
Medicare Part A and B spending ($/patient/month)	845	944	845	0.72	0.00	860[g]
30-day unplanned hospital readmission (#/person/quarter)	10.93	5.23	11.86	-0.94	-0.09	n.a.
Age as original reason for Medicare entitlement (%)	74.7	73.3	75.0	-0.23	-0.02	83.3[h]
Disability as original reason for Medicare entitlement (%)	25.2	26.5	24.9	0.22	0.02	16.7[h]
ESRD as original reason for Medicare entitlement (%)	0.5	0.7	0.5	0.00	0.03	0.1[h]

Table II.C.2 *(continued)*

Characteristic of practice	Treatment practices (N = 87)	Unmatched comparison pool (N = 7,380)	Matched comparison group (N = 290)	Absolute difference[a]	Standardized difference[b]	Medicare FFS national average
Percent of the practice's patients who are dually eligible for Medicaid	19.7	22.3	19.2	0.49	0.03	21.7[i]
Age (years)	71.0	71.0	71.3	-0.35	-0.09	71[j]
Age younger than 65 (%)	17.9	18.8	17.0	0.82	0.07	16.7[h]
Ages 65–74 (%)	42.9	41.5	42.9	0.07	0.01	45.5[h]
Age 75–84 (%)	27.5	27.2	28.0	-0.53	-0.08	25.4[h]
Age 85 or older (%)	11.7	12.6	12.1	-0.37	-0.06	12.4[h]
Female (%)	59.3	58.8	58.6	0.66	0.12	54.7[h]
Characteristics of high-risk patients attributed to practices during the baseline year (January 1, 2012 – December 31, 2012)						
Number of high-risk beneficiaries	229	101	248	-18.89	-0.10	n.a.
HCC risk score	2.32	2.36	2.35	-0.03	-0.12	1.0
All-cause inpatient admissions (#/1,000 patients/quarter)	172.67	190.64	174.08	-1.41	-0.03	74[e]
Outpatient ED visit rate (#/1,000 patients/quarter)	235.43	241.94	230.40	5.03	0.06	105[f]
Medicare Part A and B spending ($/patient/month)	1,738	1,882	1,730	7.98	0.01	860[g]
Characteristics of the practices (nonhealth centers only)						
Meaningful use of EHR (%)	93.5	45.9	90.4	3.12	0.10	n.a.
Ownership: owned by hospital or health system (%)	84.4	22.8	80.0	4.42	0.10	n.a.
Number of clinicians at practice [k]	6.6	3.1	7.4	-0.82	0.13	n.a.
Has 1 clinician (%)	2.6	48.8	3.9	-1.30	-0.06	n.a
Has 2 or 3 clinicians (%)	18.2	31.7	21.4	-3.20	-0.07	n.a.
Has 4 or 5 clinicians (%)	18.2	10.5	16.8	1.41	0.04	n.a.
Has 6 to 14 clinicians (%)	52.0	7.9	50.1	1.84	0.04	n.a.
Has 15 or more clinicians (%)	9.1	1.1	7.8	1.26	0.05	n.a.
Percentage of practices' clinicians with primary care specialty	89.8	92.2	89.4	0.42	0.02	n.a.
Omnibus tests for balance on matching variables[l]						
Health centers and nonhealth centers (combined)						
p-value			0.89			
Nonhealth centers only						
p-value			0.75			

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code household income data merged from the American Community Survey ZIP Code Characteristics.

Notes: The comparison group means are weighted based on the number of matched comparisons per treatment beneficiary. For example, if four comparison practices are matched to one treatment practice, each of the four comparison practices has a matching weight of 0.25.

Table II.C.2 *(continued)*

High-risk beneficiaries are defined as beneficiaries with a Hierarchical Condition Category score in the top quarter among all beneficiaries seen by treatment practices during the period (baseline or intervention), by market area.

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the matched treatment and comparison groups.

[b] The standardized difference is the difference in means between the matched treatment and comparison groups divided by the standard deviation of the variable, which is pooled across the matched treatment and selected comparison groups.

[c] Variables for which we required treatment and comparison members to match on exactly. For example, a treatment practice that was a health center could be matched only to a comparison practice that was a health center, and each treatment practice could match only to comparison practices in the same market area.

[d] Variables that we matched on through a propensity score, which captures the relationship between a practice's characteristics and its likelihood of being in the treatment group.

[e] Health Indicators Warehouse (2014b).

[f] Gerhardt et al. (2014).

[g] Boards of Trustees (2013).

[h] Chronic Conditions Warehouse (2014a, Table A.1).

[i] Health Indicators Warehouse (2014c).

[j] Health Indicators Warehouse (2014a).

[k] Clinicians include physicians, nurse practitioners, and physician assistants.

[l] Results from an overall chi-squared test indicate the likelihood of observing a set of differences on the matching variables that is as large as what was observed if the treatment and comparison beneficiaries in the matched sample were equivalent on all the matching characteristics indicated. For example, the value of $p = 0.89$ for the chi-squared test for the combined group (health centers and non-centers) suggests that the two groups are well balanced, because we cannot reject the null hypothesis that they are the same. We report the results of the chi-squared test separately for the nonhealth centers because, due to available data, practice characteristic variables were available for this group but not for the health centers; the nonhealth center test includes these additional variables.

*/**/*** Significantly different from zero at the .10/.05/.01 levels, two-tailed test, respectively. (Note: The primary tests assume a one-tailed test, for the reasons explained in the text.)

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; EHR = electronic health record; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = Hierarchical Condition Category.

NA = not available.
n.a. = not applicable.

Characteristics of the practices overall. Our analysis includes 87 treatment practices at the start of the intervention, ten of which are FQHCs or RHCs. Practice characteristics were available only for the non-health centers. Most treatment practices were owned by a hospital or health system (84 percent) and almost all treatment practices had providers receiving payment from the Centers for Medicare & Medicaid Services (CMS) as using EHRs in a meaningful way (94 percent). This latter proportion is consistent with TransforMED's targeting, as one of the program's eligibility criteria was an EHR system that had been actively used among practice staff for at least a year. In contrast, only about one-quarter of the practices in the pool of more than 7,000 potential comparison practices were owned by a hospital or health system (23 percent), and fewer than half were using EHR in a meaningful way (46 percent). Treatment practices, on average, had 6.6 total providers and a vast majority of providers in these practices had primary care as their specialty (90 percent).

Characteristics of the practices' Medicare FFS beneficiaries. The characteristics of all Medicare FFS beneficiaries assigned to the treatment practices during the baseline period

(January 1, 2012, through December 31, 2012) were, overall, largely similar to the nationwide FFS averages. The HCC risk score for the treatment group was close with the national average (1.1 versus 1.0). Participants in the treatment practices also had hospital admission rates and total Medicare spending that were close to the national averages. The mean outpatient ED visit rate (138/1,000 people/quarter) was higher than the national average of 105.

The high-risk beneficiaries in the treatment group had substantially greater health care needs during the baseline period than the full treatment group. The mean HCC risk score in this group was more than twice the mean for all treatment group members (2.3 versus 1.1), consistent with how the group was defined. Further, members of the high-risk group had approximately twice the number of all-cause inpatient admissions and Medicare spending and 70 percent more outpatient ED visits. These comparisons are between the high-risk subgroup and the full treatment group; differences would be even larger if we compared the high-risk group to its complement (that is, only those people in the treatment group who were not also in the high-risk subgroup).

4. Equivalence of the treatment and comparison groups at the start of the intervention

Demonstrating that the treatment and comparison groups were similar at the start of the intervention is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying difference-in-differences models—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment practices not received the intervention.

Table II.C.2 shows that the 87 treatment practices and the 290 selected comparison practices were similar at the start of the intervention on variables used in matching. By construction, there were no differences between the two groups on the market area in which practices were located. There were some differences between the treatment group and matched comparison group beneficiaries on the variables we matched through propensity scores, but the standardized differences across the propensity-score matching variables were all within our target of 0.25 standardized differences, and all were actually within 0.15 standardized differences (the 0.25 target is an industry standard; see Institute of Education Sciences 2014). The omnibus test that the treatment and comparison practices are perfectly matched on all variables common across both health and nonhealth centers cannot be rejected ($p = 0.89$), further supporting that the treatment and comparison groups were similar at the start of the intervention. Similarly, omnibus tests for the health ($p = 0.32$; results not shown) and nonhealth centers ($p = 0.75$) subgroups were not statistically significant at the $p < 0.15$ level.

The propensity-matching technique improved or did not affect the balance for most variables, relative to not matching. This can be seen in Table II.C.2, which shows the means for the full comparison pool and for the selected comparison group. Specifically, propensity matching improved balance on the number of attributed beneficiaries, the number of attributed high-risk beneficiaries, and the 30-day unplanned hospital readmission rate for all patients. Propensity matching also improved balance on several important characteristics of the practice (nonhealth centers only) including the number of providers, whether the practice was owned by a hospital or health system, and whether the practice had providers receiving payment from CMS

for using EHR in a meaningful way. Balance for some variables (for example, percentage female) was slightly negatively affected by our propensity-score matching; however, because even after matching the differences between the two groups were small in magnitude and fell well within our target of 0.25 standardized differences, we considered these slight differences to be inconsequential.

5. Intervention impacts

In this section, we first present sample sizes and mean outcomes, by quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates; however, differences in mean outcomes are not impact estimates by themselves. Next, we present the results of the primary tests (which are regression-adjusted). Then, we present the results of the secondary test results and assess whether the primary test results are plausible given the secondary results, and whether primary test results are plausible given the implementation evidence. We end with preliminary conclusions about program impacts on service use.

a. Sample sizes

Full population. In the first baseline quarter (B1), the treatment group included 79,872 Medicare FFS beneficiaries assigned to the 87 treatment practices and the comparison group included 267,072 Medicare FFS beneficiaries assigned to the 290 comparison practices (Table II.C.3.a). The size of the treatment and comparison groups grew from each quarter to the next during the baseline period. As expected, the sum of the comparison group weights was roughly equal to the size of the treatment group in each baseline quarter.

In the first intervention quarter (I1), the treatment group included 86,314 beneficiaries and the comparison group included 280,045 beneficiaries. The sample size in I1, 86,314, was in line with our anticipated sample size of 85,000 per quarter, based on information from TransforMED. The treatment and comparison samples generally grew over time during the intervention period, as in the baseline period. During the intervention period, the sum of the comparison group weights was slightly greater than the size of the treatment group in each intervention quarter, suggesting that the comparison practices grew at a slightly faster rate than the treatment practices.

High-risk subgroup. In B1, the treatment group included 20,126 high-risk Medicare FFS beneficiaries and the comparison group included 68,838 high-risk beneficiaries (Table II.C.3.b). By construction, these high-risk groups are one-quarter the size of all Medicare FFS beneficiaries assigned to the treatment and comparison panels (Table II.C.3.a). The sample sizes stayed relatively steady across the four baseline quarters for both the intervention and comparison groups. This stability reflects the net result of two opposing forces—beneficiaries being added to the sample as new high-risk beneficiaries are assigned to the treatment and comparison practices (which will increase the sample sizes) and beneficiaries exiting the sample (which will decrease sample sizes) due to death or enrolling in a health maintenance organization (and therefore becoming unobservable in FFS claims data). As expected, the sum of the comparison group weights was roughly equal to the size of the treatment group in each baseline quarter.

Table II.C.3.a. Sample sizes and unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and comparison groups for TransforMED, by quarter

	Number of Medicare FFS beneficiaries (practices)			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)		
Q	T	C (no wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (January 1, 2012 – December 31, 2012)									
B1	79,872 (87)	267,072 (290)	78,401	83.0	87.0	-4.0 (-4.5%)	131.1	138.3	-7.3 (-5.2%)
B2	83,064 (87)	276,865 (290)	82,484	78.5	78.2	0.3 (0.3%)	134.3	138.1	-3.8 (-2.7%)
B3	85,839 (87)	285,236 (290)	86,146	77.3	76.4	0.9 (1.2%)	140.6	146.6	-6.0 (-4.1%)
B4	88,568 (87)	292,928 (290)	90,312	81.0	77.8	3.2 (4.2%)	134.6	139.4	-4.8 (-3.4%)
Intervention period (January 1, 2013 – December 31, 2014)									
I1	86,314 (87)	280,045 (290)	93,027	79.3	80.8	-1.5 (-1.8%)	125.3	142.4	-17.1 (-12.0%)
I2	89,730 (87)	288,526 (290)	96,211	78.4	75.5	2.9 (3.8%)	133.5	146.2	-12.7 (-8.7%)
I3	92,675 (87)	295,641 (290)	98,967	74.5	79.8	-5.2 (-6.6%)	134.2	152.4	-18.2 (-11.9%)
I4	95,433 (87)	301,636 (290)	100,955	74.1	71.5	2.6 (3.6%)	128.5	139.3	-10.8 (-7.8%)
I5	94,815 (87)	298,569 (290)	99,602	78.4	79.1	-0.7 (-0.9%)	127.2	135.4	-8.2 (-6.1%)
I6	96,670 (87)	303,214 (290)	100,901	76.0	89.1	-13.0 (-14.6%)	134.2	144.4	-10.2 (-7.1%)
I7	98,213 (87)	307,083 (290)	102,607	72.0	79.9	-7.9 (-9.9%)	138.3	154.2	-15.9 (-10.3%)
I8	99,698 (87)	310,685 (290)	104,172	76.3	81.8	-5.5 (-6.7%)	131.2	146.5	-15.3 (-10.5%)

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The baseline quarters are measured relative to when the baseline period began on January 1, 2012. For example, the first baseline quarter (B1) runs from January 1, 2012, to March 31, 2012. The intervention quarters are measured relative to the start of the intervention period on January 1, 2013. For example, the first intervention quarter (I1) runs from January 1, 2013, to March 31, 2013.

In each period (baseline or intervention), the treatment group in each quarter includes all beneficiaries attributed to a treatment practice by the start of the quarter and enrolled in FFS Medicare. In each period, the comparison group in each quarter includes all beneficiaries attributed to a comparison practice by the start of the quarter and who met the other sample criteria. See text for details.

The outcome means were weighted, such that (1) each treatment beneficiary gets a weight of 1; and (2) each comparison beneficiary gets a weight that is the product of two weights: (a) a matching weight, equal to the reciprocal of the total number of comparison practices matched to the same treatment practice as the beneficiary's assigned practice; and (b) a practice size weight, which equals the average number of beneficiaries assigned to the matched treatment practice during the four baseline quarters divided by the average number of beneficiaries assigned to the beneficiary's comparison practice over those quarters. The

Table II.C.3.a *(continued)*

difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; Diff = difference; ED = emergency department; FFS = fee-for-service; I = intervention; Q = quarter; T = treatment; wgt = weights.

NA = not available.

n.a. = not applicable.

Table II.C.3.b. High-risk subgroup sample sizes and unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and comparison groups for TransforMED, by quarter

	Number of Medicare FFS beneficiaries (panels)			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)		
Q	T	C (uw)	C (w)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (January 1, 2012 – December 31, 2012)									
B1	20,126 (87)	68,838 (289)	19,573	189.4	202.5	-13.1 (-6.5%)	234.9	240.9	-5.9 (-2.5%)
B2	20,009 (87)	68,239 (290)	19,817	168.2	169.4	-1.3 (-0.7%)	237.4	235.8	1.6 (0.7%)
B3	19,837 (87)	67,607 (290)	19,763	161.6	162.8	-1.2 (-0.7%)	243.2	253.9	-10.7 (-4.2%)
B4	19,617 (87)	66,838 (290)	20,435	175.0	161.0	13.9 (8.6%)	237.0	244.1	-7.1 (-2.9%)
Intervention period (January 1, 2013 – December 31, 2014)									
I1	21,858 (87)	70,899 (290)	25,467	184.7	171.8	12.9 (7.5%)	227.2	247.7	-20.5 (-8.3%)
I2	21,638 (87)	69,756 (290)	25,589	174.9	152.9	22.0 (14.4%)	240.8	255.9	-15.1 (-5.9%)
I3	21,482 (87)	68,707 (290)	25,447	164.0	162.4	1.6 (1.0%)	244.0	261.4	-17.4 (-6.7%)
I4	21,288 (87)	67,491 (290)	25,034	167.1	147.8	19.3 (13.0%)	232.8	232.8	-0.0 (-0.0%)
I5	20,476 (87)	64,724 (290)	23,972	173.3	167.6	5.7 (3.4%)	227.4	222.7	4.7 (2.1%)
I6	20,100 (87)	63,127 (290)	23,133	168.8	208.1	-39.3 (-18.9%)	238.6	244.0	-5.5 (-2.2%)
I7	19,647 (87)	61,676 (290)	22,541	160.4	175.5	-15.1 (-8.6%)	249.9	257.8	-7.9 (-3.1%)
I8	19,204 (87)	60,261 (290)	21,975	174.1	177.7	-3.6 (-2.0%)	234.3	252.1	-17.7 (-7.0%)

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The baseline quarters are measured relative to when the baseline period began on January 1, 2012. For example, the first baseline quarter (B1) runs from January 1, 2012, to March 31, 2012. The intervention quarters are measured relative to the start of the intervention period on January 1, 2013. For example, the first intervention quarter (I1) runs from January 1, 2013, to March 31, 2013.

High-risk beneficiaries are defined as beneficiaries with a Hierarchical Condition Category score in the top quarter among all beneficiaries seen by treatment practices during the period (baseline or intervention), by market area.

In each period (baseline or intervention), the high-risk subgroup of the treatment group in each quarter includes high-risk beneficiaries attributed to a treatment practice by the start of the quarter and enrolled in FFS Medicare. In each period, the high-risk subgroup of the comparison group in each quarter includes all high-risk beneficiaries attributed to a comparison panel by the start of the quarter and who met the other sample criteria. See text for details.

Table II.C.3.b *(continued)*

The outcome means were weighted, such that (1) each treatment beneficiary gets a weight of 1; and (2) each comparison beneficiary gets a weight that is the product of two weights: (a) a matching weight, equal to the reciprocal of the total number of comparison practices matched to the same treatment practice as the beneficiary's assigned practice; and (b) a practice size weight, which equals the average number of beneficiaries assigned to the matched treatment practice during the four baseline quarters divided by the average number of beneficiaries assigned to the beneficiary's comparison practice over those quarters. The difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; Diff = difference; ED = emergency department; FFS = fee-for-service; I = intervention; Q = quarter; T = treatment; uw = unweighted; w = weighted.

NA = not available.

n.a. = not applicable.

In I1, the treatment group included 21,858 high-risk beneficiaries and the comparison group included 70,899 high-risk beneficiaries. This slight increase relative to the baseline period reflects (1) the total number of beneficiaries assigned to the treatment and comparison panels grew slightly between the baseline and intervention periods (for example, due to growth in the practices' Medicare patient population); and (2) at the start of the intervention period, we reset the high-risk group as those with HCC scores in the top quarter of the distribution at the start of the intervention period. The sample sizes for the treatment and comparison groups gradually declined across each of the eight intervention quarters, as more high-risk beneficiaries exited the sample than were assigned to the treatment and comparison practices.

b. Mean outcomes for the treatment and comparison groups, by quarter

Table II.C.3.a shows unadjusted treatment group means during four baseline quarters and eight intervention quarters across the two core outcomes used in the impact analysis for the full Medicare FFS population; Table II.C.3.b shows the same statistics for the high-risk subgroup of the Medicare FFS population.

All-cause inpatient admissions. The hospitalization rates for the full population during the baseline period ranged from 77.3 to 83.0 per 1,000 beneficiaries per quarter for the treatment group and 76.4 to 87.0 for the comparison group. Rates for the treatment group tended to be lower (ranging from 72.0 to 79.3) during the intervention period than the baseline period. Rates for the comparison group during the intervention period were generally higher than those for the treatment group, ranging from 71.5 to 89.1, with sizable differences between the two during the last three quarters of the period (differing by 6.7 to 14.6 percent).

For the high-risk population, hospitalization rates were generally similar between treatment and comparison beneficiaries during the baseline period; however, rates for the comparison group were consistently lower than the treatment group's during the first five intervention quarters (by 1.0 to 14.4 percent), then consistently higher than the treatment group during the last three intervention quarters.

Outpatient ED visit rates. For the full sample treatment group, the outpatient ED visit rate ranged from 131.1 to 140.6 per 1,000 beneficiaries per quarter during the four baseline quarters.

The rates were slightly lower for the treatment group during the intervention period than the baseline period, ranging from 125 to 138. The rates of the comparison group were modestly higher than the treatment group's in the baseline quarters (by 2.7 to 5.2 percent); the difference in rates between the treatment and comparison groups was larger (by 6.1 to 11.9 percent) during the eight intervention quarters.

For the high-risk population, the treatment group's outpatient ED visit rate fluctuated across quarters (from 227.2 to 249.9) but was neither increasing nor decreasing over the baseline and intervention periods. The ED visit rates were modestly higher (2.2 to 8.3 percent higher) for the comparison group than the treatment group in all intervention quarters, with the exception of I4 and I5.

c. Results for primary tests on service use

Overview. The primary test results reflect the average impact of the intervention on service use in the seventh and eighth interventions quarters (I7 and I8). These results indicate favorable effects for the full Medicare FFS population that are substantively large and statistically significant (Table II.C.4). As described previously, these results are preliminary because the analyses do not yet cover the full time period that we will include in the final impact analysis in future reports (I7 to I10).

Full Medicare FFS population. The treatment group averaged 74.1 all-cause inpatient admissions per 1,000 beneficiaries per quarter over I7 and I8, which was estimated to be 5.7 admissions fewer than the counterfactual. (Our estimate of the counterfactual is the treatment group mean minus the difference-in-differences estimate). This favorable difference between the treatment group mean and the counterfactual was statistically significant at 10 percent ($p = 0.08$, after adjusting for multiple statistical tests in the domain), and also larger than the substantive threshold (7.1 versus 5.0 percent). Similarly, we observe a favorable difference that is larger than the substantive threshold for the rate of outpatient ED visits. The treatment group averaged 135 outpatient ED visits per 1,000 beneficiaries per quarter over the time period, which was estimated to be 8.4 visits fewer than the counterfactual. As with inpatient admissions, the difference between the treatment group mean and the counterfactual was statistically significant ($p = 0.06$, after adjusting for multiple statistical tests in the domain).

High-risk subgroup. During I7 and I8, the high-risk treatment group averaged 167 inpatient admissions per 1,000 beneficiaries per quarter and 242 outpatient ED visits per 1,000 beneficiaries per quarter, which were estimated to be 5.7 admissions fewer and 7.7 ED visits fewer than the counterfactual. However, these differences were not statistically significant after adjusting for multiple statistical tests in the domain ($p = 0.25$ for inpatient admissions and $p = 0.48$ for ED visits). Although the analyses were well powered to detect impacts the size of the substantive threshold (15 percent), the observed differences for inpatient admissions and ED visits over the two quarters were considerably smaller than this threshold. Table II.C.4 shows that, if the program had succeeded in reducing inpatient admissions or ED visits by 15 percent, our analysis would have had an 82 or 83 percent probability, respectively, of detecting the effect (using a one-tailed test, and a $p < 0.10$ threshold).

Table II.C.4. Results of primary tests for TransforMED

Primary test definition					Statistical power to detect an effect that is[b]		Results			
Domain (# of tests in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage relative to the counterfactual)[a]	Size of the substantive threshold	Twice the size of the substantive threshold[c]	Treatment group mean	Regression-adjusted difference between the treatment and estimated counterfactual (standard error) [a]	Percentage difference[d]	*p*-value[e]
Service use (4)	All-cause inpatient admissions (#/1,000 beneficiaries/ quarter)	Intervention quarters 7 and 8	All observable Medicare FFS beneficiaries attributed to treatment practices	-5.0%	52.0	91.7	74.1	-5.7 (3.0)	-7.1%	0.083 [f]
	Outpatient ED visits (#/1,000 beneficiaries /quarter)			-5.0%	66.5	98.4	134.8	-8.4 (4.2)	-5.9%	0.064 [f]
	All-cause inpatient admissions (#/1,000 beneficiaries/ quarter)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices	-15.0%	81.5	99.9	167.3	-15.8 (12.6)	-8.6%	0.247 [f]
	Outpatient ED visits (#/1,000 beneficiaries/ quarter)			-15.0%	82.9	99.9	242.1	-7.7 (16.8)	-3.1%	0.479 [f]
	Combined (%)			-10.0%	94.5	100.0	n.a.	n.a.	-6.2%	0.038 [g]

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at Centers for Medicare & Medicaid Services.

Notes: The results for each outcome are based on a difference-in-differences regression model. For each intervention quarter, the model calculates the regression-adjusted difference between outcomes for the treatment and comparison groups in that quarter, subtracting out any differences between the treatment and comparison groups during the baseline period.

High-risk beneficiaries are defined as beneficiaries with a Hierarchical Condition Category score in the top quarter among all beneficiaries seen by treatment practices during the period (baseline or intervention), by market area.

[a] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[b] The power calculation is based on actual standard errors from analysis. For example, in the first row, a 5 percent effect on all-cause inpatient admissions (from the counterfactual of 74.1 + 5.7 = 79.8) would be a change of 4.0 all-cause inpatient admissions per 1,000 beneficiaries/quarter. Given the standard error of 3.0 from the regression model, we would be able to detect a statistically significant result 52.0 percent of the time if the impact was truly 4.0 percentage points, assuming a one-sided statistical test at the $p = 0.10$ significance level.

Table II.C.4 *(continued)*

[c] We show statistical power to detect a very large effect (twice the size of the substantive threshold) because this provides additional information about the likelihood that we will find effects if the program is indeed effective. If power to detect effects is less than 75 percent even for a very large effect, then the evaluation is extremely poorly powered for that outcome.

[d] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison group, divided by the adjusted comparison group mean.

[e] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test). Because it is a one-sided test, as the difference-in-differences estimate approaches positive infinity, the *p*-value approaches 1, whereas it would approach 0 in a two-sided test.

[f] We adjusted the *p*-values from the primary test results for the multiple (four) comparisons made within the service use domain.

[g] This *p*-value tests the null hypothesis that the difference-in-differences estimates across the four outcomes in the domain, each expressed as percentage change from the estimated counterfactual, is less than or equal to zero (a one-sided test).

ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award.

Combined across all primary tests. After combining results across the four primary tests in the service use domain (two outcomes across two populations), service use outcomes for the treatment group during I7 and I8 were 6.2 percent lower than the counterfactual. This favorable difference between the treatment group mean and the counterfactual was statistically significant ($p = 0.04$).

d. Results for secondary tests

As shown in Table II.C.5, the differences in inpatient admissions per 1,000 beneficiaries per quarter between the treatment beneficiaries and their counterfactual—for both the full population treatment group and the high-risk population treatment group—were positive, small (1.0 and 2.7 percent), and not statistically significant in the secondary test period (I1 through I4). These results help support the credibility of the comparison group because we do not see large differences (favorable or unfavorable) during the first year of practice participation, a period during which we and the awardee did not expect to see large program effects (see Section II.C.2.g). This increased confidence in the comparison group, in turn, gives us greater confidence in the primary test results.

e. Consistency of quantitative estimates with implementation findings

The primary test results were also plausible given implementation findings. The implementation findings suggest that TransforMED was successful in implementing the PCMN program, which involved providing practice management tools to participating practices so they could more effectively use data to improve clinical processes in a larger effort of providing better care to their patients (see Section II.A). Overall, 77 of the 90 participating practices fully completed the implementation process; 13 practices did not implement Phytel but were able to implement the cost management reporting software. TransforMED reported that it reached its program process and service goals; practices were able to meet goals related to patient contact measures and process measures (such as number of screenings and number of care plans), with several of these measures retired as communities reached 90 to 100 percent of their patient panels.

Table II.C.5. Results of secondary tests for TransforMED

Secondary test definition				Results			
Domain	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between treatment and comparison groups (standard error)	Percentage difference[a]	*p*-value[b]
Service use	All-cause inpatient admissions (#/1,000 beneficiaries/ quarter)	Intervention quarters 1–4	All observable Medicare FFS beneficiaries attributed to treatment practices	76.7	0.7 (2.3)	1.0%	0.624
	All-cause inpatient admissions (#/1,000 beneficiaries/ quarter)		All observable high-risk Medicare FFS beneficiaries attributed to treatment practices	172.7	4.5 (8.6)	2.7%	0.700

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at the Centers for Medicare & Medicaid Services.

Notes: The results for each outcome are based on a difference-in-differences regression model, as described in the text.

High-risk beneficiaries are defined as beneficiaries with a Hierarchical Condition Category score in the top quarter among all beneficiaries seen by treatment practices during the period (baseline or intervention), by market area.

[a] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison groups, divided by the adjusted comparison group mean.

[b] The *p*-values from the secondary test results were *not* adjusted for multiple comparisons within the domain.

FFS = fee-for-service.

f. Conclusions about program impacts on service use

Based on the evidence available, we have concluded that the program had a statistically significant favorable effect on service use during the first two quarters of the primary test period (Table II.C.6). The primary tests for all-cause inpatient admissions and ED visits among the full Medicare FFS population were favorable and statistically significant (after adjusting for four tests in domain); the combined test for the service use domain was favorable and statistically significant; the secondary tests confirmed the plausibility of the primary tests; and implementation findings indicate it is plausible that TransforMED's PCMN intervention was implemented in a manner that could have affected service use.

As mentioned earlier, these conclusions are preliminary because the analyses do not yet cover the full period nor all the outcomes that we will include in the final impact analysis in future reports.

Table II.C.6. Preliminary conclusions about the impacts of TransforMED's HCIA program on participants' outcomes, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion		
		Primary test result(s) that supported conclusion	Primary test result plausible given secondary tests?	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	n.a	n.a	n.a	n.a
Service use	Statistically significant favorable effect	Estimate for all-cause inpatient admissions was favorable and statistically significant (after adjusting for four tests in domain) Estimate for outpatient ED visits was favorable and statistically significant (after adjusting for four tests in domain) Estimate for combined estimates (4) in the service use domain was favorable and statistically significant	Yes	Yes
Service use	n.a	n.a	n.a	n.a

Sources: Tables II.C.4 and II.C.5.

ED = emergency department; HCIA = Health Care Innovation Award.

n.a. = not applicable.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

TransforMED received HCIA funding to assist 14 participating health systems and 90 primary care practices to implement population management systems and cost-management reporting. TransforMED also facilitated collaboration within and between the communities and participating practices through shared learning activities. The program aimed to improve performance on condition-specific quality measures and patients' experiences of care, and reduce the overall costs of care for Medicare and Medicaid beneficiaries. The awardee also planned to expand the program to additional practices in each participating community. TransforMED experienced delays in program launch and implementation of the population management and cost-management reporting software systems, Phytel and Cobalt Talon. Despite these delays, overall TransforMED implemented the HCIA-funded program largely as intended. Program implementation was facilitated by the availability of cost and quality data, the development of new relationships between providers in the PCMN communities, and its alignment with other health care delivery initiatives. In addition, the implementation climate created by health system leadership as well as health systems' and practices' external technology environments (that is, the formation of a state health information exchange [HIE]) were facilitators in one of the two

communities visited. Challenges related to dedicating resources to program implementation, mapping the new software to existing EHRs, and engaging stakeholders across the PCMN communities hindered implementation. Payment models and lack of reimbursement for adopting PCMH concepts were challenges in one of the communities we visited. The HCIA Primary Care Redesign Clinician Survey found that almost half of clinicians reported a lack of familiarity with the TransforMED program, but clinicians' views generally aligned with the goals of PCR, and most clinicians rated most of the PCR goals as extremely important.

The impact evaluation found favorable and statistically significant impacts of the program on service use (all-cause hospitalizations or outpatient ED visits) for Medicare FFS beneficiaries during the first 6 months of the primary test period for these outcomes (months 19 through 24 after the program began). The impact on service use was driven by large impact estimates for hospitalizations and ED visits among the full population of Medicare FFS beneficiaries that exceeded the substantive threshold of 5 percent, although we found no measurable effects on service use outcomes for the high-risk subgroup of Medicare FFS beneficiaries.

Our next steps for this evaluation are to (1) monitor TransforMED's program implementation reports through June 30, 2015, and plans for sustaining the program beyond the funding period; (2) evaluate trainees' and clinicians' attributes and experiences with the program in the third year of the award through administered surveys; (3) extend the impact evaluation to include the full period of program operations and Medicare Part A and B spending as an outcome; and (4) use the implementation findings to help interpret the impact results.

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

REFERENCES

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table A.1. Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

Chronic Conditions Data Warehouse. "Table B.2. Medicare Beneficiary Prevalence for Chronic Conditions for 2003 Through 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014b. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_b2.pdf. Accessed November 19, 2014.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014c. Available at http://www.healthindicators.gov/Indicators/Medicare-beneficiaries-eligible-for-Medicaid-percent_317/Profile/ClassicData. Accessed August 4, 2015.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." Rockville, MD: Agency for Healthcare Research and Quality, AHRQ Publication No.11-0100-EF, October 2011.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for University Hospitals of Cleveland Rainbow Babies and Children's Hospital

March 2016

Joseph Zickafoose
Brenda Natzke
Boyd Gilman
Catherine DesRoches
Lorenzo Moreno

Submitted to:
U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day

Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno

Reference Number: 40274.270

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

CONTENTS

TABLES

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

UNIVERSITY HOSPITALS OF CLEVELAND RAINBOW BABIES AND CHILDREN'S HOSPITAL

This individual program report summarizes the findings to date from our evaluation of the primary care redesign (PCR) program implemented by University Hospitals of Cleveland Rainbow Babies & Children's Hospital (UHC) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A). We then describe the attitudes and behavior of the clinicians affected by the program (Section II.B). The third evaluation component—estimating the impact of the program on patients' outcomes—is not available for this awardee due to delays in obtaining Medicaid data. In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF UHC

UHC received a three-year, $12.8 million HCIA to transform the delivery of health care for children enrolled in Medicaid in northeastern Ohio. The UHC program aimed to create a sustainable pediatric ambulatory care model that improves care and health and lowers costs for children enrolled in Medicaid by creating arrangements with pediatric primary care providers, and relationships with patients and their caregivers, Medicaid managed care organizations (MCOs), and the state Medicaid agency. Table I.1 summarizes key features of the program. Its goals (core outcomes) include reducing avoidable emergency department (ED) visits by 15 percent for Medicaid-enrolled children, reducing the total cost of care by 2.5 percent for Medicaid-enrolled children, having 75 percent of participating pediatricians meet quality targets, and enrolling at least 750 children with complex chronic or behavioral health conditions in care coordination services. Although the program aimed to improve care for Medicaid-enrolled children, all interventions were open to all children regardless of insurer. UHC received a no-cost extension to continue program activities through March 2016.

Table I.1. Summary of UHC PCR program

Awardee name	University Hospitals of Cleveland Rainbow Babies & Children's Hospital
Award amount	$12,774,935
Implementation date	January 2013
Award end date	March 2016
Program description	1. Establish structural components to support the functioning of the clinical components and develop a sustainable financial model, including • A primary care provider network for care improvement activities • Shared savings contracts with Medicaid MCOs • A program database to support population health and care gap analyses 2. Establish six core clinical components to improve care for children, including • Practice facilitation for primary care practices • Care coordination for children with complex chronic conditions • Integrated behavioral health services • ED avoidance interventions • Patient and community outreach for alternatives to ED care • Hospital readmission prevention

Table I.1 *(continued)*

Innovation components	Care coordination, care management, care transitions, health IT, integrated team care, payment reform, workforce development, practice facilitation, home care, and telehealth
Intervention focus	Individual and practice
Workforce development	Create new positions (practice facilitators, behavioral health social workers, care managers, care coordinators, community health workers, and telehealth physicians and attendants); change roles and responsibilities of existing staff (primary care providers and office staff)
Target population	Medicaid-enrolled children, children with chronic conditions, children with behavioral and mental health disorders, and frequent users of ED services
Program setting	Provider-based (hospital and primary care practices), both employed and independent practices
Market area	Local (eight counties in northeastern Ohio)
Market location	Urban, suburban, and rural
Core outcomes	• 15.0 percent reduction in avoidable hospital ED visits • 2.5 percent reduction in total cost of care • 75.0 percent of primary care providers meet quality targets

Source: Review of UHC program reports, March 2015.

Note: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health IT systems, and undertaking other operational activities

II. SUMMARY OF FINDINGS

A. Program implementation

This section first provides a detailed description of the intervention, highlighting adaptations over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related measures. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of UHC's program implementation on a review of the awardee's quarterly reports and self-monitoring program measures, telephone discussions and follow-up communications with program administrators, and information collecting during site visits conducted in April 2014 and March 2015. We did not attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

UHC's HCIA program included three structural and six clinical components. The structural components of the program included (1) engagement of a primary care provider network for care improvement activities, (2) shared savings contracts with Medicaid MCOs, and (3) a programmatic database to support population health and care gap analyses. First, UHC recruited a network of employed and contracted independent primary care practices to participate in practice facilitation and paid semiannual incentive payments based on achievement of targeted quality performance measures. At the beginning of the award, UHC established an advisory

council composed of providers from the network of practices. Throughout the award, the council provided feedback on the clinical components of the program, particularly practice facilitation and the quality improvement goals. Second, to help sustain the program, UHC set a goal of establishing contractual arrangements, including shared savings, with all five Medicaid MCOs in Ohio. Third, the program sought to establish a system for analyzing Medicaid claims to assess population health care use and costs for children enrolled in Medicaid managed care and attributed to participating practices.

The clinical components of the program included (1) primary care practice facilitation in quality measurement and improvement, (2) clinical care and care coordination for children with complex chronic conditions, (3) integrated behavioral health services, (4) ED avoidance interventions, (5) patient and community outreach for alternatives to ED care, and (6) hospital readmission prevention. We briefly describe each of these clinical program components in Table II.A.1.

The practice facilitation intervention was established for the network of practices, began by targeting improvement in three quality measures, and expanded the number of measures over time. Table II.A.2 lists and defines the quality measures by year of introduction in the program.

Several program components experienced adaptations from what was originally planned. For example, the multidisciplinary complex care component was designed to provide care coordination and direct care to children with complex medical conditions. This component also included a pilot of an on-demand home video connectivity program between the patient and the care team through tablet devices for high-risk children and those who live long distances from the complex care clinic, which was not expanded beyond a pilot phase due to challenges with Internet connectivity for families. Likewise, one of the ED avoidance components, the community-based telehealth hubs, was under consideration for discontinuation due to low use. As of March 2015, UHC administrators were exploring possibilities for on-call program physicians to staff telehealth hubs being installed in the region by retail pharmacies rather than having hospital-owned telehealth hubs. The after-hours clinic was an urgent care clinic started in December 2014 in response to lower-than-expected use of the telehealth hubs.

UHC adapted many of its case management and community outreach programs primarily in support of the ED avoidance interventions. The nurse case management program began targeting frequent ED utilizers and expanded to include infants younger than age 1 with any ED visit in an effort to prevent an ongoing pattern of avoidable ED use. In late 2014, the program began sharing lists of children with frequent or nonurgent ED use with primary care practices for targeted outreach and care coordination activites by providers and other practice staff. The program staff also provided education on ED alternatives to the families of children identified as frequent ED utilizers while they were in the UHC pediatric ED. UHC expanded these education efforts to include new mothers during their postpartum stays in the UHC maternity hospital. To support its program outreach and education efforts, UHC trained and began using a group of community health workers in spring 2014 to engage families about where and when to seek care for their children.

Table II.A.1. Key details about clinical program design and adaptation

Program component	Target population	Patient identification	Patient recruitment and enrollment	Service delivery protocol	Adaptations
Primary care practice facilitation	Employed and contracted independent primary care practices serving children enrolled in Medicaid in the region who were recruited through telephone, email, and in-person contacts; program sought to recruit enough practices or providers to reach 65,000 Medicaid-enrolled children	Not applicable; all patients at participating practices were included	No active patient recruitment or enrollment	Based on a prior RCT;[a] facilitators visited network practices weekly, performed a small number of chart audits to assess performance on quality measures, shared measure results with providers and staff, and discussed opportunities for improvement; monthly audits on larger numbers of charts were performed to assess progress toward quality targets and eligibility for incentive payments	Yes; expanded number of practices in network; combined facilitator and chart reviewer role; decreased frequency of visits to every other week for practices that were performing well; and increased number of measures (3 in the first year, 7 in the second year, 11 in the third year)
Clinical care and care coordination for children with complex chronic conditions	Children who (1) have significant neurocognitive impairment, (2) have three or more body systems impaired, (3) are technology–dependent, or (4) require caregiver assistance with activities of daily living	Review of internal administrative and billing databases	Program staff contacted caregivers of eligible children by telephone and scheduled an initial appointment with the multidisciplinary team for those who agreed to participate	Initial multidisciplinary clinical evaluation with medical, nursing, social, and nutritional assessments; standardized portable care plan; worked with child's existing primary and specialty care providers; tiering system based on medical, nutritional, and social needs to determine the frequency of family contacts, ranging from every six months for less complex cases to every two months for more complex patients	Yes; formalized identification and enrollment processes; dedicated staff to contact and schedule new patients; continually revised tiering system to improve it
Integrated behavioral health services • Referral service • Telephone consultation • In-office behavioral health evaluation ED crisis intervention	Children who receive care at network practices or in UHC pediatric ED for behavioral health needs	Provider referral or identification from ED triage system	Child's caregiver contacted by telephone (referral service, in-office behavioral health evaluation) or in person (ED crisis intervention)	Referral service helped providers link children to appropriate community-based mental health agencies; behavioral health social workers available for telephone consultation to providers and initial evaluations in the child's primary care practice; crisis intervention social workers in the ED assessed patients and linked them to appropriate resources	Yes; ED crisis intervention expanded from ages 15 years and younger to 17 and younger in 2015

Table II.A.1 *(continued)*

Program component	Target population	Patient identification	Patient recruitment and enrollment	Service delivery protocol	Adaptations
Integrated behavioral health services • Referral service • Telephone consultation • In-office behavioral health evaluation • ED crisis intervention	Children who receive care at network practices or in UHC pediatric ED for behavioral health needs	Provider referral or identification from ED triage system	Child's caregiver contacted by telephone (referral service, in-office behavioral health evaluation) or in person (ED crisis intervention)	Referral service helped providers link children to appropriate community-based mental health agencies; behavioral health social workers available for telephone consultation to providers and initial evaluations in the child's primary care practice; crisis intervention social workers in the ED assessed patients and linked them to appropriate resources	Yes; ED crisis intervention expanded from ages 15 years and younger to 17 and younger in 2015
ED avoidance interventions • Telephone triage • Telehealth hubs • After-hours clinic	Children receiving care in network practices (telephone triage); any child younger than 18 (telehealth hubs, after-hours clinic)	Not applicable; this program serves all children who either present at telehealth hub or after-hours clinic or call practice after hours	Not applicable	Telephone triage protocol to assess the diagnosis, provide advice for management, and recommend future care; physician consultation for all children triaged to the ED; nurse authority to prescribe prescriptions for low-acuity conditions	No
				Community-based telehealth hubs available outside regular office hours for urgent care, staffed through video conferencing with an on-call pediatrician, supported by an in-person medical attendant	Yes, staffed by medical assistants or emergency medical technicians, one hub located in a community center and a second in a storefront space
				After-hours urgent care clinic on main UHC campus	Yes, clinic developed to meet goal of decreased ED use after low use of telehealth hubs

Table II.A.1 *(continued)*

Program component	Target population	Patient identification	Patient recruitment and enrollment	Service delivery protocol	Adaptations
Patient and community outreach for alternatives to ED care • ED use case management	Children with frequent ED use (four or more visits in 12 months), or any use before age 1 year	Used internal billing data to identify children with four or more visits to a UHC health system-affiliated ED in the prior 12 months (or any use before age 1 year); several Medicaid MCOs also provided lists of frequent utilizers	Nurse case managers contacted caregivers by telephone	Nurse case managers contacted families within two days of their ED visit and subsequently at 2, 6, and 12 weeks after the ED visit to identify ongoing concerns, barriers to care, and potential solutions	Yes; began with frequent ED utilizers and expanded to infants younger than 1 year; in late 2014, the program began sharing lists of children with frequent or nonurgent ED use with primary care practices for targeted outreach by providers and practice staff
Patient and community outreach for alternatives to ED care	Child caregivers in neighborhoods with high rates of ED use	Not applicable; target communities identified by geographic analysis of frequent utilizers of UHC pediatric ED	Not applicable	Variety of outreach campaigns, including billboards, bus advertisements, and computerized interactive calls	In spring of 2014, trained and began using a group of community health workers to engage families about where and when to seek care for their children
Hospital readmission prevention	All pediatric patients hospitalized in medical and surgical units at UHC except those from hematology/oncology or intensive care units	Patients admitted to medical and surgical units	Not applicable	Unit staff notified the primary care provider or home specialty provider upon admission and discharge of a child, ensured a follow-up appointment was scheduled before discharge, and conducted a follow-up call after discharge; facilitator reviewed charts to determine units' performance on measures, provided feedback through a unit scorecard, and brainstormed with unit leadership about quality improvement	No

Source: Interviews from second site visit, March 2015; document review, March 2015.

[a] Meropol, S.B., A. Sattar, K.C. Stange, A.H. Nevar, C. Davey, G.A. Ferretti, D.E. Howell, R. Strosaker, P. Vavrek, S. Bader, M.C. Ruhe, and L. Cuttler. "Practice-Tailored Facilitation to Improve Pediatric Preventive Care Delivery: A Randomized Trial." *Pediatrics,* vol. 133, no. 6, 2014, pp. e1664–e1675.

RCT = randomized controlled trial.

Table II.A.2. Quality measures for primary care practices participating in the UHC HCIA program

Quality measure	Definition	Year 1	Year 2	Year 3
Fluoride varnish application	Apply every 6 months after tooth eruption (12 to fewer than 36 months)	X	X	X
Obesity	Calculate BMI and percentile, diagnose weight, counsel	X	X	X
Lead screening	Order appropriate testing at 12 and 24 months	X	X	X
Asthma management	Assess control of asthma using a standardized measure		X	X
Drug formulary	Prescribe following drug formulary		X	X
URI	Do not use antibiotics for URI diagnosis		X	X
Well-child care ages 3 to 6	Schedule patients ages 3 to 6 years for follow-up well visit at time of sick visit, if needed		X	X
ADHD	Document follow-up care for patients ages 6 to younger than 18 years with a diagnosis of ADHD and prescription for a stimulant medication			X
Adolescent vaccine (T-dap/meningococcal/HPV)	Document age-appropriate doses of meningococcal, tetanus, diphtheria, acellular pertussis (Tdap), and human papillomavirus (HPV) vaccines			X
Pharyngitis	Provide appropriate evaluation and treatment for patients ages 2 to younger than 19 years diagnosed with pharyngitis			X
Well-child care ages 13 to <18	Schedule patients ages 13 to younger than 18 years for follow-up well visit at time of sick visit, if needed			X

Source: Int Interviews from second site visit, March 2015; document review, March 2015.

ADHD = attention deficit hyperactivity disorder; BMI = body mass index; HPV = human papilloma virus; URI = upper respiratory infection.

The hospital readmissions program began in August 2014 and was built on previous hospital efforts with the addition of a quality improvement facilitator, modeled on the practice facilitation component

b. Target populations and patient identification, recruitment, and enrollment

Table II.A.1 provides key details about the target populations and the patient identification, recruitment, and enrollment processes for each component. Although the overall target population for the program was children enrolled in Medicaid, the program components varied in their specific target populations and recruitment and enrollment strategies, and the interventions were open to all children regardless of insurer.

c. Service delivery protocols

Table II.A.1 also provides key details about the service delivery protocols for each component of the program. The practice facilitation component was based on an intervention developed for a prior randomized controlled trial but, for the remaining components, staff developed and adapted new protocols. For example, the complex care team developed a qualitative tiering system based on an assessment of a patient's medical, nutritional, and social complexity and used this system to determine the frequency of family contacts, ranging from two to six months depending on the complexity of the patient's case.

d. Intervention staff and workforce development

Each component of the UHC program required creating new staff positions and responsibilities. At the administrative level, a medical director designed interventions, recruited and managed relationships with network practices, provided advice and feedback to staff, and served as the public face of the program. Initially, a senior program director managed the day-to-day operations of the program and staff, established staff and managerial roles, and created program databases and information feedback mechanisms to staff. Together, the medical director and program director collaborated with the UHC health system legal and managed care departments to negotiate with Medicaid MCOs. After the program director left the program in June 2014, the medical director, the manager of the practice facilitation and ED avoidance components, and the manager of the telephone triage program shared these responsibilities. The manager of the practice facilitation component had master's-level training and had previously served as a practice facilitator. The director of the UHC health system-affiliated Medicare and commercial accountable care organizations (ACOs) also assumed some of the responsibilities for managing the practice network, managed care contracts, and population health data. A data analyst managed the program databases and data reports.

Table II.A.3 describes frontline staff members, their responsibilities, and role adaptations. Most components had no staff adaptations with some exceptions. For example, the complex care component also initially trained medical assistants to serve as care navigators for families. However, the project team realized that medical assistants did not function well in this role due to the complexity of enrolled patients' needs. UHC discontinued this portion of the intervention and used the resources to increase professional staff time in the complex care coordination team, including nursing, social work, and nutrition.

Program training occurred primarily through standard hospital personnel training sessions, on-the-job shadowing of staff with the same or similar roles, and on an ad hoc basis as staff identified additional useful skills and knowledge. For example, a social worker in the integrated behavioral health services component attended a certificate program for integrated behavioral health services in primary care, and the practice facilitation manager began a quality improvement training program through Cincinnati Children's Hospital.

Table II.A.3. Key details about intervention staff and workforce development

Program component	Staff members	Staff /team responsibilities	Adaptations?
Primary care practice facilitation	Practice facilitators	Worked directly with network practice staff to perform chart review for quality measures, review results, discuss quality improvement, and address questions related to the program; master's degrees in public health or epidemiology or significant experience in research or quality improvement settings	Yes; role merged with chart reviewer in mid-2014 following cross-training of staff to perform both roles
	Chart reviewers	Reviewed a larger number of patients' charts on a monthly basis to assess practices' quality measures for incentive payments; typically had bachelor's degrees and significant experience in research or similar settings	Yes; role merged with practice facilitator in mid-2014 following cross-training of staff to perform both roles
Care coordination for children with complex chronic conditions	Medical director	Designed interventions and oversaw operations; provider in the complex care clinic	No
	Nurse manager	Designed interventions, oversaw operations, and directed quality improvement	No
	Nurse practitioners	Served as providers in the complex care clinic and, later, in the hospital as consultants for children in the complex care program who had been admitted	No
	Nurse care coordinators	Participated in intake evaluation and needs assessment; interacted with patients via telephone, by meeting families at scheduled primary or specialty care visits, and through home visits to families at scheduled intervals and as needed to help manage family needs and crises; helped families set goals, develop self-management skills, and coordinate care among health care providers and other support services in the child's life, such as human services agencies and schools; managed children in the program with most active or complex medical needs	No
	Social work care coordinators	Participated in intake evaluation and needs assessment, interacted with patients via telephone, by meeting families at scheduled primary or specialty care visits, and through home visits to families at scheduled intervals and as needed to help manage family needs and crises; helped families set goals, develop self-management skills, and coordinate care among health care providers and other support services in the child's life, such as human services agencies and schools; managed children in the program with most complex social needs	No
	Dieticians	Performed an initial comprehensive nutritional evaluation and provided ongoing follow-up by telephone depending on the acuity of the child's nutritional needs	No
Integrated behavioral health services	Office-based behavioral health social workers	Provided telephone consultation to providers and families, helped providers link patients to community-based integrated mental health agencies, and worked in network primary care offices to provide on-site mental health evaluations	Yes, positions initially divided between telephone-based services and on-site work in primary care practices, but modified roles to cover all services
	ED-based crisis intervention social workers	Evaluated children presenting to the ED with acute behavioral health needs, staffed their evaluations with pediatric psychiatrists, and referred and connected families to appropriate sites of care, including hospitalization or outpatient community mental health agencies	Yes, adapted hours to cover times of greater demand and when not performing behavioral health evaluations, provided education to caregivers of children with nonurgent ED visits

Table II.A.3 *(continued)*

Program component	Staff members	Staff /team responsibilities	Adaptations?
ED avoidance intervention	Telephone triage nurses	Employed by an existing telephone triage service at UHC; practices maintained separate contracts with UHC for the nurse telephone triage services, and for patients enrolled in these program network practices triage nurses were able to prescribe treatment for low-acuity conditions (for example, diaper rash or ringworm) and were expected to have the on-call physician speak with the family of any child prior to being referred to the ED	No
	On-call physicians	Performed third and final level of triage for children who might need referral to the ED as indicated by the triage nurse; provided direct telehealth care through the telehealth hubs.	No
	Medical attendants	Performed intake for patients to the telehealth hub, assisted physician exams using electronic exam equipment (for example, a stethoscope or an otoscope), performed basic testing such as rapid streptococcal throat swabs or urinalysis, and provided patients with final information; required to be certified as a medical assistant or emergency medical technician	No
	Nurse practitioners	Served as providers in the after-hours clinic	No
Patient and community outreach for alternatives to ED care	Nurse case managers	Bachelor's-level nurses who contacted families of children with frequent or nonurgent ED visits to discuss diagnosis and treatment, access to appropriate follow-up care, and options for care other than the ED	No
	Community outreach coordinator	Master's-level staff member with training in health education and promotion who coordinated marketing campaigns, collaborated with community organizations, and worked with the community health workers to distribute information about assessing illness in children, connecting with primary care, and seeking care in setting other than the ED	No
	Community health workers	Members of the communities targeted by the ED avoidance interventions who had at least a high school diploma; performed community outreach through door-to-door contacts and community events to educate caregivers on the importance of a primary care visit and seeking care for illness in settings other than the ED	No
Hospital readmission prevention	Readmissions facilitator	Reviewed charts to determine the hospital units' performance on discharge quality measures, provided feedback via a unit scorecard, and brainstormed with unit leadership about approaches to improving performance; master's degree in public health or epidemiology or significant experience in research or quality improvement settings	No
	Medical director	Designed intervention and oversaw operations	No

Source: Interviews from second site visit, March 2015; document review, March 2015.

2. Implementation effectiveness

This section examines the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, timeliness, and selected service- and staff-related measures, relying on self-reported information included in UHC's quarterly self-monitoring and measurement reports.

a. Program enrollment

The effectiveness of program enrollment varied by component. In the primary indirect service delivery component, primary care practice facilitation, UHC exceeded its target enrollment of practices (Table II.A.4). In the direct service delivery components, enrollment in the complex care component fell short of the original target, but enrollment in the integrated behavioral health services component exceeded it. Program staff reported that the original target for complex care underestimated the time needed to complete the initial evaluations and conduct ongoing follow-up with children and their families. The ED avoidance and family and community outreach components did not have specific enrollment goals. However, program staff reported that telephone triage use and nurse case management enrollment met their expectations, but the use of the telehealth hub was lower than expected.

Table II.A.4. Targeted versus actual program enrollment

Program component	Targeted enrollment	Actual enrollment	Above or below target
Primary care practice facilitation	28 practices providing care to more than 65,000 children in Medicaid managed care	32 practices providing care to more than 71,000 children in Medicaid managed care	Above
Clinical care and care coordination for children with complex chronic conditions	500	178	Below
Integrated behavioral health services	650	More than 3,600 through psychiatric social workers; more than 560 through ED crisis social workers	Above

Source: Interviews from second site visit, March 2015; document review, March 2015.

b. Program time line

After six months of planning and infrastructure development, and shortly after CMMI's approval of the project plan, UHC began delivering program services to eligible children and participating practices in January 2013. The activities proceeded mostly according to schedule, although negotiating shared-savings agreements with Medicaid MCOs and opening the telehealth hubs were both delayed. UHC was able to negotiate shared-shavings agreements with two MCOs close to its original schedule, but agreements with two other MCOs took about one year longer and no agreement was reached with the fifth MCO. UHC originally planned to have both telehealth hubs fully operational by the first quarter of 2013, but the first hub began operating

only in late 2013 and the second in early 2015. The availability of leasable space in the target neighborhoods that could meet medical facility certification standards delayed the second hub significantly. Creating a claims-based population health database, needed to guide most program activities across components, also experienced delays due to obtaining data from the state and Medicaid MCOs and challenges with the validity and usability of the data once received. The vendor meant to create the population health database terminated the contract with UHC because it felt it could not provide the contracted deliverables using the available data. UHC contracted with a new vendor and provided it with updated data from the state.

c. Service measures

UHC was largely successful in reaching its broad program process and service delivery goals. As of March 2015, the program had recruited a provider network composed of 32 practices, with 164 pediatric providers across 51 locations providing care to more than 71,000 children enrolled in Medicaid managed care, exceeding its goal of 65,000 children. Of these practices, about 60 percent were employed by subsidiaries of UHC's parent health system and 40 percent were independent. At the same time, UHC executed shared-savings contracts with four Medicaid MCOs and received shared-savings payments from all four during the award period. UHC received monthly data from two MCOs, enabling it to identify patients for several of the clinical components of the program, particularly the ED avoidance interventions.

UHC had mixed experience in establishing databases to support program activities and population health data analysis. The program established an operational database that staff in all components of the program used to track program processes, plan workflows, and identify areas of improvements. However, throughout the award, challenges with the integrity of data feeds caused the program to struggle to create a population health database using the state's Medicaid claims data. After an experienced national vendor was unable to implement the planned population health database, UHC switched in spring 2015 to another vendor used by its adult ACOs.

In its self-measurement and monitoring plans, UHC reported mixed results for service delivery measures across each component and did not have explicit goals or benchmarks for most measures. First, the multidisciplinary care team for the complex care intervention contacted the families of more than 90 percent of the enrolled children within the scheduled time frame during most of the award period. However, the proportion of families of enrolled children contacted within the scheduled time frame fell below 80 percent in early 2015. Second, there was less use of the telehealth hubs than expected, with no more than 45 visits in any month and fewer than 30 visits in most months of operation. Third, in the family and community outreach component, the team reported several hundred contacts per month with families by the outreach coordinator, the community health workers, and automated telephone calls. Fourth, within the first several months of the hospital readmission program, UHC reported that more than 70 percent of eligible patients had a completed discharge summary within 24 hours of discharge, a scheduled follow-up appointment before discharge, a follow-up telephone call within 48 hours of discharge, and an admission notification letter copied to the primary care provider within 24 hours of discharge. Finally, UHC assessed parents' satisfaction with four of the program

components (behavioral health, complex care, telehealth hub, and telephone triage) and reported consistently high satisfaction rates (more than 90 percent) throughout the award period.

d. Staffing measures

UHC met its goals for new hires into the program. Cumulatively, it hired 53.75 new full-time equivalents, compared with a goal of 53.13, but was at about only 80 percent of projected staffing in the spring of 2015 due to staff attrition and combining staff roles. The newly hired staff were spread across the various components of the program, including administration (20 percent), primary care practice facilitation (20 percent), care coordination for children with complex chronic conditions component (15 percent), integrated behavioral health services (14 percent), ED avoidance and alternatives (30 percent), and hospital readmission prevention (2 percent). When UHC lost staff, it either hired new staff or redistributed responsibilities among existing staff. Notable changes in staffing occurred in the practice facilitation and integrated behavioral health components. The practice facilitation component began with 8.0 full-time staff members, which was reduced to 6.5 by March 2015 after the consolidation of the practice facilitator and chart reviewer roles. The integrated behavioral health component increased staff from two psychiatric social workers and one ED crisis intervention social worker to three psychiatric social workers and two ED crisis intervention social workers due to the increased demand for services.

3. Implementation experience

This section uses the Consolidated Framework for Implementation Research (CFIR) to analyze implementation experience and the internal and external factors associated with implementation effectiveness. We review three domains associated with implementation experience: (1) implementation process, (2) program characteristics, and (3) internal and external environments. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.5 summarizes the major facilitators and barriers to UHC's implementation effectiveness in each domain.

Table II.A.5. Facilitators and barriers to implementation effectiveness

Domain	Facilitators	Barriers
Program characteristics	• User control • Adaptability • Relative advantage	No significant barriers noted
Implementation process	• Self-monitoring and quality improvement • Stakeholder engagement • Program resources	Provider engagement Quality of data needed to monitor progress
Internal factors	• Team characteristics • Health information and other technology • Organizational culture and leadership	Health information and other technology Limited prior history
External factors	• Patients' needs and resources • General policy environment	Patients' needs and resources General policy environment

Source: Interviews from second site visit, March 2015; document review, March 2015.

a. Program characteristics

Three characteristics of the UHC program made the most substantial contributions to its ongoing implementation: (1) frontline users' flexibility in implementing the program (user control), (2) the adaptability of the program to meet patients' and providers' needs, and (3) providers' perceptions of the relative advantage of the program compared with the standard delivery of care. First, frontline staff described flexibility to modify their roles and workflows to meet the needs of the program and its patients. For example, program administrators originally planned to have one social worker staff the telephone referral line and another provide behavioral health evaluations in primary care offices. The two people hired for these roles identified logistical challenges with this arrangement and modified their roles to share both telephone and primary care office-based work. As another example, in the complex care component, the care team originally used a tiering system based primarily on medical need. The team members quickly noted that this did not capture the variety of other needs of a child, such as social or nutritional needs, so the team changed the tiering system to reflect these other needs. During the March 2015 site visit, the complex care team members also noted how they modified their roles over time to meet the specific needs of families: "I've been to the shelters to help with placement to help some of the families that are homeless. One of the social workers who is not here anymore, she assisted with a sibling's college applications. So there's really not a lot that we won't do to help the family out. We're not just helping the patient; we really have to stabilize the entire family."

Second, the administrators have adapted the program to fit the community and practices' contexts and achieve the broader goals of the program. Multiple administrators and staff noted that the program aimed to be innovative and that changes to the operational plan were expected and necessary for effective implementation. Although the core components of the program remained unchanged, administrators added subcomponents, such as hiring community health workers to expand the reach of the ED avoidance component and adding ED case managers to decrease unnecessary ED visits. The ED case management program adapted over time to include caregivers of children younger than 1 year with any ED visit in addition to the original focus on children with four or more ED visits in the prior 12 months. The program also made other changes to improve efficiency, such as merging the practice facilitator and chart reviewer roles and decreasing the amount of data collected for the practice facilitation component compared with the original research-based protocols.

Third, program administrators, providers, and primary care office staff felt the program offered advantages over the status quo, especially for behavioral health. Respondents noted challenges before the award in having sufficient time to address behavioral health concerns in office visits and obtaining access to behavioral health services for patients. As one social worker noted in the March 2015 site visit, "Before pediatricians were handing families a list or had a couple names and families would call and were not able to get an appointment.… We have time to focus on those families and make those calls for them and make sure there is availability … Really kind of helping to ease them into it because it can be a very frustrating process when your

kid is struggling with depression … and then they're making a call and feeling like they are going around in circles." Several providers felt that the program supported them in being more consistent in meeting measures for high quality care. A few providers and staff also mentioned the importance of the program in reducing the cost of care.

b. Implementation process

Three implementation process factors significantly facilitated the implementation of the UHC program: (1) self-monitoring and quality improvement, (2) stakeholder engagement, and (3) program resources. First, program staff and administrators established a data infrastructure for the program that facilitated implementation by enabling them to track workflows, measure intermediate outcomes, and begin to assess proxies for core outcomes, such as costs of care. Administrators and staff established an internal program database to track key workflow processes and intermediate outcomes, such as patient and family functionality. For example, the behavioral health staff can use the database to track incoming referrals, note when families make follow-up appointments, and monitor family functioning scores at designated intervals after referral. They also use data to tailor services to the times of highest demand. As one administrator noted about the ED crisis social work intervention during the March 2015 site visit, "I'm looking at the current staffing model for that social worker, it is from 2:00 pm to midnight, and I'm looking at when the bulk of patients are coming in. We did a demand analysis and realized that if we did 11:00 a.m. to 11:00 p.m., we could go from capturing 50 percent to 75 percent of patients in the ED." In addition, UHC used data to support the practice facilitation and family and community outreach components of the program. Staff in the practice facilitation component used a combination of weekly small-sample and monthly larger-sample chart reviews to assess quality measures. The weekly reviews also provided frequent feedback to practices on performance to enable ongoing quality improvement; the monthly reviews provided more robust measures of performance to feed back to practices and served as the basis for distribution of semiannual incentive payments to practices. The family and community outreach components used internal hospital and Medicaid MCO data to identify children with frequent or nonurgent ED visits to help target nurse case management and general family education efforts. This information also helped to identify neighborhoods with high rates of ED use to determine where to place the telehealth hubs.

Second, program leadership and staff engaged a broad group of stakeholders to support program implementation. The practice facilitation component engaged practice leaders and staff through initial meetings and continued engagement through direct practice facilitation, a quality-based incentive plan, provider advisory group meetings, and continuing medical education events. UHC used the facilitation model in the readmission avoidance component to engage leaders and staff in hospital medical and surgical units. UHC successfully engaged four Medicaid MCOs in shared-savings agreements and is pursuing a more sophisticated value-based payment contract with all five Medicaid MCOs in the state. The program has also successfully involved a large group of community mental health agencies to connect families to the agencies' behavioral health interventions.

Third, program and network practice staff and providers noted that the resources available, both through the award and as in-kind donations from UHC, promoted program implementation

(note, we do not have data on the level of in-kind resources provided to the program). Staff and providers at all the practices visited during site visits noted that the practice facilitators were easy to work with, provided valuable information, and were responsive to requests. As part of a large health care system, the program also had access to in-kind resources outside the award, including a human resources department for hiring, legal and contracting departments for managed care negotiations, contracting, and regulatory guidance, information technology support, and a health system ACO infrastructure. The program also had access to a variety of clinical experts to provide consultation on specific topics, such as asthma care quality improvement.

UHC encountered two notable implementation process challenges: provider engagement and the quality of the data needed to monitor progress. First, although the program exceeded its targeted number of practices and providers for its network, program administrators noted that some providers were slow to embrace efforts to reduce unnecessary utilization and costs. For example, several providers were reluctant to implement changes that might reduce the number of fee-for-service visits, such as allowing telephone triage nurses to call in prescriptions for low-risk acute conditions, because they were concerned about a negative financial impact on their practices due to decreased volume of patients with these common conditions. Program administrators tried to address these concerns by educating concerned providers about the health and other ancillary benefits to their patients in this model through the medical advisory council and individual interactions with providers. In addition, ongoing positive interactions with practice facilitators helped providers identify more closely with the goals of the program and engage with the interventions. As one provider in the March 2015 site visit noted, "I think overall it has been a good program. We're definitely better off than where we were in measuring quality measures. It was a journey and not always pleasant. Just in figuring out what they wanted from us. It was helpful to remember what the point of this was." Second, the Medicaid data obtained from the state vendor (and needed to monitor utilization and cost trends) had missing fields and other errors that took time to identify and correct. The data also lacked cost information and the program had to work with a partner of the database vendor to create proxy costs from utilization data. The program is now pursuing an independent, third-party evaluation of the program to help UHC understand the effects of the interventions on costs.

c. Internal factors

Characteristics of the organization implementing a program and features of the environmental contexts in which the organization is located can also influence implementation. Three internal factors positively influenced implementation of the UHC program: (1) team characteristics, (2) health information and other technology, and (3) organizational culture and leadership.

First, the structure and functioning of teams facilitated implementation within and across components. The program leaders considered the role of teamwork during hiring throughout the award, emphasizing hiring people who would be proactive and work well in teams. Shared roles and responsibilities, ad hoc peer consultation, and formal team meetings all contributed to strong team functioning within components, especially practice facilitation, integrated behavioral health, and complex care, and for the overall program. In the spring of 2014, nearly all of the program staff (except the complex care staff) relocated to a single physical location after

previously working in separate locations. Staff from multiple components described how this facilitated collaboration across components and improved understanding of the overall goals of the program. Practice facilitators in particular described how relocation enabled them to assist practices with understanding and accessing all program components.

Second, health information and other technology served as facilitators to some aspects of the program. Program staff and administrators were able to leverage UHC's health system experience with data analytics in commercial ACOs to develop data systems and use internal billing data to inform decisions about the program. Although practices were often in many different stages of electronic health record (EHR) adoption and used different products, the practice facilitators found it relatively easy to adapt to different systems to collect data through chart reviews. In the case of at least one independent practice, staff assisted with customization of their EHR to adjust their workflow and structured data collection to promote their ability to meet quality goals. In the case of the telehealth hubs, on-call physicians and program administrators felt that after a trial period, the technology worked well to deliver care virtually.

Third, organizational and program leadership and culture facilitated implementation in many ways. UHC health system administrators expressed a commitment to organizational innovation and moving toward more value-based delivery and payment models, including commercial and Medicaid ACOs. Within the program, staff noted that the administrators fostered a collaborative environment and were available to assist with problems, but also encouraged staff to take ownership and responsibility of tasks. Program administrators promoted a culture of continuous quality improvement in which staff described frequently looking for ways to improve their work through the use of new information and small changes.

Two internal factors served as an implementation barrier: health information and other technology, and prior history. Several staff members, especially in the complex care component, described the challenges of trying to use an EHR for care coordination and quality improvement when it was not designed with these functions in mind. Because the EHR did not function well in this regard, the complex care staff used a paper-based system for care coordination. They also described the structural challenges to the health system's EHR, particularly that inpatient and outpatient systems were not linked.

Limited prior experience implementing similar interventions among UHC administrators and staff was also a barrier to program implementation. Although the primary care practice facilitation model was built upon a prior research trial, the other components of the program were built from the ground up. As a result, staff underestimated the time needed for complex care evaluations, the challenges of identifying and contracting for locations for the telehealth hubs, and limited acceptance of the telehealth hubs.

d. External environment

Two external factors have facilitated implementation: (1) patients' needs and resources and (2) the general policy environment. First, patients' needs and resources were taken into account during the design and implementation stages of the program. For example, the locations of the comprehensive care clinic and telehealth hub were based on analyses of patients' zip codes and

community resources. The practice facilitation included measures (such as fluoride application and lead testing) that addressed issues for which the population is at high risk. Second, the organization was well connected to external organizations, such as Medicaid MCOs, community behavioral health agencies, and providers and agencies that care for populations with complex chronic conditions. UHC leveraged these connections, some of which predated the award, to facilitate behavioral health referrals and identify locations for its telehealth hubs. Second, the general policy environment facilitated several components of the initiative. For example, the Ohio legislature passed legislation before the beginning of the HCIA program encouraging the formation of Medicaid pediatric ACOs (though implementing regulations have not been developed); in early 2015, regulations requiring Medicaid to pay for telehealth services became effective. In addition, the Ohio Medicaid agency encouraged program administrators to include behavioral health in the program.

The same two external factors also served as barriers to implementation: (1) patients' needs and resources and (2) the general policy environment. First, program administrators and staff noted that, by definition, children enrolled in Medicaid generally live in situations of significant poverty, presenting additional barriers to improving care—such as lack of transportation, unstable housing, and lack of social supports—and these barrier often compound one another. For example, the complex care component attempted to overcome challenges with transportation for families through a pilot using tablet computers to allow for on-demand video conferencing with the care team. However, most families lacked a reliable Internet connection to support the use of the devices. Second, the Medicaid policy environment in Ohio presented challenges. Ohio Medicaid rebid its managed care contracts in 2013, causing delays in UHC's ability to negotiate contractual arrangements with MCOs. At the time of the award, the state did not require MCOs to engage in value-based agreements, so negotiations were often slow.

4. Sustainability and scalability

UHC program leaders actively planned for sustainability from the beginning of the program, and they intended for the program to continue after the award. They implemented shared-savings agreements with four Medicaid MCOs and had begun discussions with all five MCOs in the state about moving beyond shared savings to other more comprehensive value-based contracts. The current shared-savings agreements were based on ED visits, pharmacy, and quality, and program leaders hope to move to contracts that include total cost of care, quality, and a care coordination fee. However, progress has been slow and they were unsure if the arrangements would be in place by the end of the award. They had also begun to explore possibilities for value-based contracts with commercial insurers that could use and support the same infrastructure. In addition, they engaged state Medicaid officials and broader hospital leadership in discussions on value-based payment. Program leaders expected that University Hospitals' existing infrastructure and expertise with multiple ACOs would facilitate the sustainability of the pediatric ACO and its clinical components. At the time of this writing, UHC is preparing to submit applications for bridging funding from foundations and other donors to finance program operations after the end of the award and before longer-term value-based agreements are in place. UHC is also exploring the feasibility of obtaining direct reimbursement for several program, including the ED crisis intervention services, telehealth hub visits, after-hours clinic visits, and care coordination for

children with complex chronic conditions. One issue that might arise while UHC leaders pursue avenues for sustainability is the trade-off between providing services to all relevant patients, which is administratively simpler and was noted to be more ethically acceptable to providers, versus focusing on patients with specific insurance types, which is challenging financially without supportive contracts with multiple payers. UHC is also exploring ACO arrangements with commercial payers to help address this issue.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from March 2015 site visit interviews with program leadership and frontline staff at selected clinical sites or satellite offices provided important insights into the implementation process. However, in a large program such as the HCIA-funded PCR program implemented by UHC, these interviews present the perspectives of a relatively small number of people. Although these in-person interviews provide a rich source of data, views from the leadership and staff at a small number of clinical locations might differ from clinicians' views overall. In order to assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in the fall of 2014, the third year of the HCIA-funded program. Data from the survey provided additional insight into the implementation process and experience as well as the contextual factors that might affect implementation effectiveness at UHC.

In this section, we report on the views of clinicians in the practice-tailored facilitation component of the program regarding of their daily work life and practice. Clinicians associated only with care coordination for children with complex chronic conditions, integrated behavioral health services, ED avoidance interventions, patient and community outreach for alternatives to ED care, or hospital readmission prevention components were not surveyed. First, we focus on the contextual factors that affected program implementation, including the characteristics of the practice location, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care team functions. We then present data on the alignment of clinicians' views and experiences with the overall goals of the HCIA-funded innovation, their awareness of and participation in the program, and their views of the facilitators of and barriers to successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice locations

A total of 88 clinicians in the network of practices in the practice-tailored facilitation component of the program responded to the survey (resulting in a response rate of 64 percent). Of the respondents, 77 were physicians and 11 were nurse practitioners. (The number of clinicians in each response category throughout this section do not always sum to the total number of respondents due to survey item nonresponse, as well as clinicians who reported that a given question was not applicable to their practice and thus did not provide a response.) Nearly three-quarters of respondents (74 percent) were in a group practice with three or more clinicians, with the rest working in solo or two-clinician practices, community health centers, health

systems, health maintenance organizations, or other practice arrangements. More than half of the respondent clinicians' primary method of compensation was a salary adjusted for productivity (60 percent); another 28 percent received a fixed salary.

All clinicians participating in the UHC program who responded to the survey reported working in settings that use some form of health IT. Nearly all clinicians reported basic health IT functions (Table II.B.1), much higher than a national estimate that half of physicians practice in settings with functional EHRs (Furukawa et al. 2014). However, smaller proportions of clinicians in the UHC program reported using advanced health IT functions such as patient registries or patient-facing functions, which are not in widespread use nationally (DesRoches et al. 2014).

Table II.B.1. Health information technology functions

Survey item	Number of respondents	Percentage of respondents
Ordering tests and procedures	72	82%
Accessing laboratory test results	86	98%
Prescribing medications	81	92%
Drug dosing and interaction alerts	86	98%
Entering clinical notes	88	100%
Electronic referral tracking	53	60%
Patient registries	35	40%
Patient-facing functions:		
Request a prescription refill	40	45%
Request an appointment	28	32%
Email a clinician about a medical question or concern	37	42%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

b. How clinicians experience their careers and workdays

Clinicians participating in the UHC program were generally satisfied with their careers, with nearly all very (56 percent) or somewhat satisfied (35 percent) with their careers and most reporting not feeling burned out (82 percent). In addition, nearly two-thirds of clinicians (65 percent) reported spending 75 percent or more of their week doing work that is well matched to their training, which could explain their satisfaction levels.

Clinicians participating in the UHC program similarly rated their workplace management in a positive light. About half of respondents strongly agreed that their management team was supportive (53 percent), they were encouraged to offer suggestions and improvements (48 percent), they had adequate opportunities for professional development (52 percent), and the amount of work they were expected to complete each day was reasonable (53 percent).

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. Most clinicians somewhat or strongly agreed that

it is possible to provide high quality care to all of their patients (73 percent), but many still reported significant barriers, including time for patient care, information from other physicians, reimbursement, and patients' difficulty paying for needed care (Table II.B.2).

Table II.B.2. Perceptions of limits to providing high quality care

Survey item	Number of respondents	Percentage of respondents
Percentage reporting each of the following at least somewhat limits their ability to provide optimal, patient-centered care		
I do not have enough time to spend with patients during visits	76	86%
I lack timely information about the patients I see who have been cared for by other physicians	62	70%
The level of reimbursement is not adequate	59	67%
My patients have difficulty paying for needed care	49	56%
I receive too many reminders from my EHR	35	40%
It is difficult for me to obtain specialist referrals for my patients in a timely manner	32	36%
It is difficult for me to obtain specialized diagnostic tests or treatments for my patients in a timely manner	31	35%
I lack adequate information from research evidence to guide my clinical decisions	28	32%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

c. Clinicians' perceptions of care team functioning

In the survey, a care team was defined broadly as clinicians and other staff involved in caring for patients, such as medical assistants, care coordinators, registered nurses, and community health workers. Almost two-thirds of clinicians (65 percent) participating in the UHC program reported working as part of a care team, and they had mostly positive perceptions of these teams. For those who worked in a care team, most agreed that members of the care team relayed information in a timely manner (97 percent), had sufficient time for patients to ask questions (95 percent), used common terminology when communicating with one another (93 percent), verbally verified information they received from one another (86 percent), and followed a standardized method of sharing information when handing off patients (62 percent).

d. Alignment with goals of primary care redesign

The survey included several items asking clinicians to rate the importance of a series of goals related to achieving high quality, patient-centered care on a scale ranging from extremely important to not important at all. The inclusion of the extremely important category helped to force respondents to choose between goals that are essential to meet and those that are simply important. In Table II.B.3, we present results based on the proportion of clinicians rating each of these goals as extremely important. The views of clinicians participating in the UHC program generally aligned well with the specific goals of the UHC program, but not always with some of the overall goals of PCR. For example, more than half of the respondents indicated it is extremely important to improve care coordination for patients with chronic conditions, reduce ED visits, increase access to and improve continuity of primary care, and improve the

appropriateness of care, all of which are key goals of the UHC program. One explicit goal of the UHC program was of lower priority for clinicians; only about one-third considered it extremely important to reduce overall health care spending.

Table II.B.3. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Improving care coordination for patients with chronic conditions	53	60%
Reducing ED visits	52	59%
Increasing access to primary care	52	59%
Improving care continuity in primary care	48	55%
Improving appropriateness of care	45	51%
Reducing hospital readmissions	43	49%
Increasing the use of evidence-based practice in clinical care	41	47%
Improving patients' capacity to manage their own care	35	40%
Improving the capability of health care organizations to provide patient-centered care	34	39%
Reducing overall health care spending	31	35%
Improving capability of health care organizations to provide team-based care	27	31%
Increasing the number of primary care practices functioning as a PCMH	27	31%
Increasing use of EHRs and other health IT	17	19%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

EHR = electronic health record; IT = information technology; PCMH = patient-centered medical home.

A large majority of clinicians (92 percent) were engaged in quality improvement activities in the past two years, including collaborative efforts with other practices, hospitals, government agencies, or professional associations; training on quality improvements and tools (88 percent), or at least one clinical audit of care that their patients received (80 percent). These activities align well with the goals and activities of the UHC program for participating primary care clinicians, although we are unable to identify through the survey the degree to which the responses were based on UHC program activities or other quality improvement efforts.

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the UHC program was to create a sustainable system of interventions that improved care and health and lowered costs for children enrolled in Medicaid. Understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on patients' outcomes. For example, if clinicians are aware of the program, have received appropriate and effective training, and believe that the program will have a positive effect on the care they provide, they are likely to feel more invested in the program's success. Alternatively, those who feel more negatively about the program might be less likely to

enthusiastically implement the intervention. In this section, we report on clinicians' experiences with and perceptions of the UHC program.

a. Awareness of the program and receipt of training

Ninety-two percent of the clinicians participating in the UHC program we surveyed were at least somewhat familiar with the program. Of these clinicians, 91 percent had received training related to the program. On average, clinicians received 9.9 hours of program-related training.

b. Perceived effect of program on patients' care

Clinicians' perceptions of the effect of the UHC program on the care they provide to patients were mixed. As shown in Table II.B.4, most clinicians who were familiar with the program believed it would have a positive effect on the quality of the care they provide, their ability to respond to patients' needs in a timely way, and the patient-centeredness of care. Fewer clinicians perceived that the program would have a positive impact on efficiency, safety, and equity of care. Few clinicians (12 percent or fewer) thought the program would have a negative impact on aspects of patients' care, but a substantial minority of clinicians believed the program would have no effect on the care they provided (data not shown). A small proportion of clinicians (10 to 15 percent) reported it was too soon to tell what effects the program was having on patients' care (data not shown).

Table II.B.4. Clinician perceptions of effects of program on patient care

	Positive effect		No effect	
Survey item	Number of respondents	Percentage	Number of respondents	Percentage
Perceived effect of the HCIA program on the care they provided to patients over the past year, including on:				
Quality of care	61	75%	-	-
Ability to respond in a timely way to patients' needs	41	51%	27	33%
Efficiency	27	33%	34	42%
Safety	35	43%	36	44%
Patient-centeredness	41	51%	27	33%
Equity of care for all patients	37	46%	32	40%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Clinicians were asked about the perceived effect of the program and the barriers to and facilitators of implementation only if they reported being at least somewhat familiar with the program. Figures are based on the total number of clinicians reporting they were at least somewhat familiar with the UHC program. Cells with fewer than 11 respondents are not reported due to confidentiality restrictions.

c. Barriers and facilitators to program implementation

Finally, we asked clinicians who were at least somewhat familiar with the UHC program to rate the effect of a series of barriers and facilitators to program implementation. More than half

of the clinicians who were familiar with the program rated six factors as having a positive effect on program implementation: level of program funding (57 percent), availability of personnel (60 percent), availability of relevant patient information at the point of care (52 percent), availability of evidence-based clinical information (57 percent), availability of community resources to care for complex patients (57 percent), and quality of interpersonal communications with specialists (54 percent). The main barrier to implementing the program was the amount of required documentation, which 41 percent of clinicians reported as having a negative impact.

4. Conclusions about clinicians' attitudes and behavior

The results from the clinician survey provide evidence of an environment conducive to effective implementation of the UHC program. Surveyed clinicians were generally satisfied with their careers, did not report feeling burned out, and spent the majority of their time doing work that is well-matched to their training. They view their workplace management and care teams positively, and utilize at least some form of HIT. Nearly all clinicians were at least somewhat familiar with the UHC program, and nearly all had received training for the program and were engaged in quality improvement efforts. Clinicians viewed as important many of the overall goals of primary care redesign that align with the specific goals of the UHC program, although fewer reported reducing overall health care costs as an extremely important goal. Approximately three-quarters of clinicians familiar with the program believed it would have a positive effect on the quality of care they provide. They identified many facilitators to program implementation and few barriers. These findings are consistent with information we collected during our site visits in which practice site staff endorsed the goals of the program, described program staff and activities as responsive to their practices' needs, and describe program staff and activities as helpful in the goal of improving the quality of health care for children.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

In the original application for HCIA funding, UHC proposed to create "a sustainable pediatric ambulatory care system that improves health, improves care, and reduces costs" for children enrolled in Medicaid. Over the course of operational planning and implementation, the program expanded and adapted beyond a focus on primary care to include redesign in urgent, emergency, and inpatient care settings as a foundation for a pediatric ACO for Medicaid-enrolled children. After nearly three years of HCIA funding, UHC has implemented the components of the program largely according to plan. Implementation was facilitated by a practice facilitation program that engaged a broad network of primary care providers and a program culture of self-monitoring and quality improvement based on analyses of available data sources and feedback from frontline staff. Program implementation was hindered by a lack of usable Medicaid claims data for the target population, an underestimation of the staff time needed to provide clinical and care coordination services to children with complex chronic conditions, and challenges with engaging families with new models of care, such as the telehealth hubs. Results from the HCIA Primary Care Redesign Clinician Survey also provided evidence of effective implementation with large majorities of clinicians being familiar with and receiving training from the program, being engaged in quality improvement efforts, supporting the goals of the program, and believing the program had a positive effect on the quality of care.

Our next steps for this evaluation are to (1) monitor ongoing program implementation and plans for sustaining the program beyond the no-cost extension funding period by reviewing quarterly data submitted by UHC; (2) evaluate trainee and clinician attitudes and experiences with the program in the third year of the award through administered surveys; (3) complete agreements to obtain Ohio Medicaid data and perform an impact evaluation during the final option year of the evaluation contract; and (4) synthesize implementation, survey, and impact evaluation findings to assess the success of the UHC program in meeting its goals to improve health care for children enrolled in Medicaid in northeast Ohio.

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

REFERENCES

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no. 9, 2014, pp. 1672–1679.

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

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

VOLUME II: INDIVIDUAL PROGRAM SUMMARIES

Evaluation of the Health Care Innovation Awards (HCIA): Primary Care Redesign Programs

Second Annual Report: Findings for Wyoming Institute of Population Health at Cheyenne Regional Medical Center – Patient-Centered Medical Home Component

March 2016

Emily Ehrlich
Andrea Wysocki
KeriAnn Wells
Boyd Gilman
Greg Peterson
Catherine DesRoches
Sandi Nelson
Laura Blue
Keith Kranker
Kate Stewart
Frank Yoon
Jelena Zurovac
Lorenzo Moreno

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Project Officer: Timothy Day
Contract Number: HHSM-500-2010-000261/HHSM-500-T0015

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Lorenzo Moreno
Reference Number: 40274.270

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

CONTENTS

TABLES

FIGURES

WYOMING INSTITUTE OF POPULATION HEALTH AT CHEYENNE REGIONAL MEDICAL CENTER – PATIENT-CENTERED MEDICAL HOME COMPONENT

This individual program report provides a summary of the findings to date from our evaluation of the primary care redesign (PCR) program implemented by Wyoming Institute of Population Health (WIPH) under Health Care Innovation Award (HCIA) funding from the Center for Medicare & Medicaid Innovation (CMMI). Section I provides an overview of the WIPH program. Section II presents a summary of the evaluation findings. We first assess the effectiveness of program implementation (Section II.A) and then describe the attitudes and behaviors of the clinicians affected by the program (Section II.B). Finally, we analyze the impact of the PCMH program component on participants' outcomes (Section II.C). In Section III, we synthesize the main findings and describe the next steps of the evaluation.

I. OVERVIEW OF WIPH

WIPH received a three-year, $14.2 million dollar HCIA award to transform rural care delivery through the creation of medical neighborhoods across Wyoming. The Wyoming Medical Neighborhoods program included five components: (1) transformation of primary care practices into patient-centered medical homes (PCMHs), (2) hospital transition assistance for participants 65 or older with one of 10 qualifying conditions, (3) telehealth videoconferencing technology in hospitals and doctors' offices, (4) community-based access to free medications, and (5) the Virtual Pharmacy program. Table I.1 summarizes key features about program design. By the end of the award, in June 2015, the initiative aimed to reduce hospital emergency department (ED) visits by 10 percent, hospital admissions by 5 percent, and total spending by 5 percent.

Table I.1. Summary of WIPH's PCR program

Program feature	Wyoming Institute of Population Health
Award amount	$14,246,153
Implementation date	October 10, 2012
Award end date	June 30, 2015
Program description	Create medical neighborhoods in Wyoming that include at least one of the following: (1) patient-centered medical homes, (2) hospital transition assistance for participants 65 or older with at least one qualifying condition, (3) telehealth, (4) community-based access to free medications, and (5) the Virtual Pharmacy program
Innovation components	Patient-centered care, care transitions, telehealth videoconferencing technology in hospitals and doctors' offices, increased medication access through medication donation, medication therapy management, and care coordination
Intervention focus	Practice and participant
Workforce development	WIPH contracted with TransforMED to host quarterly learning collaboratives for transforming primary care practices; WIPH provided a 2-day training for care transition nurses
Target populations	Patients with chronic conditions, patients 65 or older, and patients with Medicaid
Program setting	Provider (hospital- and practice-based)
Market area	Statewide plus one hospital in Nebraska
Market location	Urban and rural

Table I.1 *(continued)*

Core outcomes	• 10 percent reduction in ED visits, • 5 percent reduction in hospital admissions • 5 percent reduction in total spending • Improvement in clinical outcomes, patients' engagement and satisfaction • Reduction in preventable adverse drug events • Improved access to primary care and prescription medication

Source: Review of Wyoming Institute of Population Health's program reports, March 2015.

Notes: The implementation date represents when programs began taking concrete steps toward launching their program components by hiring staff, establishing partnerships, investing in health information technology systems, and undertaking other operational activities.

WIPH is a Division of Cheyenne Regional Medical Center and acts as a convener for members of the Wyoming Integrated Care Network (WYICN). WYICN includes 23 hospital and health care facilities, along with the University Of Wyoming College Of Health Sciences and six health care professional organizations. WIPH leveraged existing strategic partnerships to build medical neighborhoods throughout the state. The awardee engaged in the following activities:

- Contracted with TransforMED, a subsidiary of the American Academy of Family Physicians that trains primary care practices in becoming medical homes, to serve as a practice facilitator supporting the PCMH transformation program
- Coordinated with participating hospitals to implement and provide leadership to nurses for the care transition program
- Assigned Cheyenne Regional Medical Center's telehealth department to lead the telehealth program
- Partnered with the Wyoming Department of Health (WDH) to coordinate the community-based Medication Donation Program (MDP)
- Partnered with the University of Wyoming's School of Pharmacy to lead the Virtual Pharmacy program

II. SUMMARY OF FINDINGS

A. Program implementation

In this section, we first provide a detailed description of the intervention, highlighting how it has been adapted over time. Second, we review the evidence of implementation effectiveness, including an assessment of measures of enrollment, implementation schedule, and other service- and staff-related metrics. Third, we examine the facilitators and barriers associated with implementation effectiveness, including those related to program characteristics, implementation processes, internal factors, and external environments. Finally, we discuss findings related to program sustainability and scalability. We based our evaluation of WIPH's program implementation on a review of the awardee's quarterly reports and self-monitoring program metrics, telephone discussions and follow-up communications with program administrators, and information collected during site visits conducted in April 2014 and April 2015. We did not

attempt to verify the quality of the performance data reported by awardees in their self-measurement and monitoring reports.

1. Program design and adaptation

a. Program components

WIPH medical neighborhoods included five component strategies. First, the WIPH's central initiative was the transformation of primary care clinics into PCMHs that formed the core of the Wyoming medical neighborhoods. The PCMH settings were diverse, including independent physician practices, hospital-based practices, rural health clinics, and Federally Qualified Health Centers (FQHCs).

Second, the Wyoming Rural Care Transitions program (WyRCT) trained hospital-based nurses to manage transitions for participants discharged from acute care settings. At two sites in Cheyenne, the awardee piloted a similar outpatient-based health coaching program, Transition across Community Teams (TACT). TACT nurses offered similar services as WyRCT nurses with the goal of preventing hospitalizations.

Third, the telehealth component provided infrastructure for provider connectivity to facilitate care coordination and increase access to care. The awardee installed desktop and mobile video conferencing technology at clinics, hospitals, and PCMHs across Wyoming.

Fourth, WIPH partnered with the WDH to lead the Wyoming MDP to increase access to medication for eligible uninsured and underinsured low-income patients.

Finally, the School of Pharmacy at the University of Wyoming (UW) led the Virtual Pharmacy program. Participating pharmacists provided participants with medication therapy management service at local pharmacies and communicated information about participants' medication use and adherence to the provider. The Centers for Medicare & Medicaid Services did not renew funding for Virtual Pharmacy for the third award year, and the awardee closed the program in July of 2014.

b. Target populations, risk assessment, recruitment, and enrollment

As described in Table II.A.1, the target populations and recruitment and enrollment strategies varied, depending on the program component. For example, the PCMH program was open to all patients regardless of insurance type or status, and it did not use any type of formal enrollment process. In contrast, the MDP targeted un- or underinsured patients who might have trouble affording their medications; providers identified and referred eligible patients to the program.

Table II.A.1. Key details about program design and adaptation

	Program component				
	PCMH	Wyoming Rural Care Transitions	Telehealth	Medication Donation	Virtual Pharmacy
Target population	• All patients regardless of insurance • Care management efforts specifically targeted patients with diabetes, hypertension, pediatric asthma, and lifestyle choices such as tobacco use	• Patients ages 65 and older, regardless of insurance status, with at least one of 10 qualifying conditions: CHF, COPD, coronary artery disease, diabetes, stroke, medical/surgical back disorder, hip fracture, peripheral vascular disease, cardiac arrhythmia, or pulmonary embolism • Participating hospitals determined an eligible service area, such as within 50 miles of the discharging hospital or within the same county as the discharging hospital	• All patients who required consultation with an outside specialist not available at the site • Patient consultations were for mental/behavioral health, bariatrics, rheumatology, endocrinology, and oncology • Usually offered to patients with Medicaid or a commercial payer offering telehealth reimbursement	• Patients with incomes up to 200 percent of the FPL • Patients with no prescription coverage • Patients on the Wyoming Prescription Drug Assistance Program who require three or more prescriptions per month • Medicare beneficiaries struggling with the Medicare Prescription Drug Plans (Part D) coverage gap	Medicaid patients ages 18 to 65 with depression/bipolar disorder, pain, asthma, cardiovascular disease, gastroesophageal reflux disorders/ulcers, or diabetes
Identification strategy	None	WyRCT nurses checked the hospital census daily for patients who met eligibility criteria	Participating physicians identified patients and available specialists	Participating physicians identified eligible patients in need of non-narcotic medications	Pharmacists reviewed patients' panel data to identify eligible patients
Recruitment/ enrollment strategy	All participating clinic patients could benefit from PCMH; there was no direct enrollment process	• WyRCT nurses approached patients during their hospital stays to explain and offer the program • Patients signed consent forms to participate	There was no direct enrollment process	Participating physicians referred eligible patients to the program	• Pharmacists enrolled eligible patients either when they came to the pharmacy to fill a prescription or via proactive outreach • Patients signed consent forms to participate

Table II.A.1 *(continued)*

	Program component				
	PCMH	**Wyoming Rural Care Transitions**	**Telehealth**	**Medication Donation**	**Virtual Pharmacy**
Service delivery protocol	• Did not follow a prescribed protocol, although practices provided team-based, patient-centered, accessible, coordinated, care • Practices offered same-day appointments and extended office hours	• Within 24 to 48 hours of discharge, WyRCT nurses visited participants at home • Following home visits, nurses called participants for 30 to 90 days; the frequency of calls varied based on the participant's needs • Nurses did not provide clinical or skilled nursing care; rather, the nurses offered services to help participants manage their own health, such as chronic condition education, home risk assessment, and medication reconciliation • Nurses also helped align participants with needed social services and attended physician visits with some participants	• Did not follow a prescribed protocol; physicians and hospitals used telehealth for live video consultations with other physicians or patients throughout the state and for trainings and administrative meetings as needed • Patients accessed remote telehealth services from local clinical sites, such as the patient's PCMH practice	• Did not follow a prescribed protocol, however WDH recruited donation sites via marketing brochures, posters, and its website • WDH solicited non-narcotic prescription drug donations via outreach to nursing homes, assisted living facilities, detention centers, and other sites that could be sources of unused medications • Licensed pharmacists disposed of unusable drug donations, such as narcotics and expired drugs • Patients access donated medications via mail or at dispensing sites	• Pharmacists provided at least 3 counseling sessions to participants • Pharmacists used motivational interviewing to help participants set goals, offered medication management therapy, and created a medication adherence plan • Pharmacists administered the SF-12 and PHQ-9 depression screening instruments • Pharmacists faxed participants' information to primary care physicians
Adaptations	Yes. In response to feedback from transforming practices, TransforMED adapted its practice facilitation approach from providing a general overview of patient-centered care delivery to assisting practices with the NCQA application process	Yes. At two sites in Cheyenne, the awardee piloted an outpatient-based health coaching program called TACT. Nurses offered similar services as WyRCT with the goal of preventing hospitalizations. TACT expanded the target population to include patients ages 18 to 65 and added qualifying conditions such as depression and anxiety. Physicians referred patients to TACT nurses	Yes. In the first quarter of 2015, WIPH implemented a home-based telehealth component, which is not generally reimbursable, for patients willing to self-pay	No	No

Source: Interviews from second site visit, April 2015; document review, March 2015.

CHF = congestive heart failure; COPD= chronic obstructive pulmonary disease; NCQA = National Committee for Quality Assurance; PCMH = patient-centered medical home; PHQ-9 = Patient Health Questionnaire; SF-12 = 12-Item Short Form Health Survey; TACT = Transition across Community Teams; WDH = Wyoming Department of Health; WIPH = Wyoming Institute of Population Health; WyRCT = Wyoming Rural Care Transitions program.

c. Service delivery protocols

WIPH detailed participant intervention protocols for WyRCT and the Virtual Pharmacy programs. PCMH, telehealth, and MDP did not have prescribed intervention protocols, although these programs provided service delivery that aligned with a patient-centered care delivery model. Table II.A.1 provides additional details about the service delivery for each program component.

d. Intervention staff and workforce development

WIPH used the HCIA funding for WyRCT and TACT nurses' salaries for the duration of the award period. WIPH allocated full-time equivalent (FTE) positions to participating hospitals based on discharge volume, and hospitals hired registered nurses (RNs) to fill the positions. WIPH also offered pharmacists participating in the Virtual Pharmacy program a capitated payment for each participant. WIPH did not use HCIA funding to compensate intervention staff participating in the other program components (Table II.A.2).

Table II.A.2. Key details about intervention staff

Staff members	Credentials	Staff/team responsibilities	Adaptations	Cumulative FTEs
WyRCT nurse	RN	• Enrolled eligible patients in the hospital • Conducted home visit within 48 hours of discharge • Provided chronic condition management coaching and education • Provided medication reconciliation • Accompanied some participants to primary care visits • Directed participants to community-based services • Collected and recorded participants' data	No	22.8
TACT nurse	RN	• Received physician referrals in outpatient setting • Conducted home visit with eligible participants • Provided chronic condition management coaching and education • Provided medication reconciliation • Directed participants to community-based services • Collected and recorded participants' data	No	2
Virtual Pharmacy pharmacist	Pharm. D.	• Conducted outreach to eligible patients • Enrolled eligible patients • Provided at least 3 counseling sessions to participants • Used motivational interviewing to help participants set goals, offered medication management therapy, and created a medication adherence plan • Administered PHQ-9 and SF-12 • Faxed participants' summaries to primary care providers	No	0.9

Source: Interviews from second site visit, April 2015; document review, March 2015.

FTE = full-time equivalent; PHQ-9 = Patient Health Questionnaire; RN = registered nurse; SF-12 = 12-Item Short Form Health Survey; TACT = Transition across Community Teams program; WyRCT = Wyoming Rural Care Transitions program.

WIPH used award funds to support various workforce development activities. WIPH contracted with TransforMED to provide leadership and practice facilitation for the PCMH transformation, as well as training and support to participating primary care sites. Training covered quality improvement activities, with an emphasis on PCR and transformation. TransforMED conducted site visits, led telephone calls, and convened quarterly learning collaboratives to support workforce development. TransforMED also helped sites develop work plans for transformation and reviewed documents for the National Committee of Quality Assurance (NCQA) PCMH recognition application.

The WyRCT and TACT programs provided nurses with two days of in-person classroom training that covered medication reconciliation, motivational interviewing, and chronic condition education and coaching.

The telehealth program did not use HCIA funding to support staff or training needs; information technology (IT) staff installed the software and led the informal internal trainings.

For MDP, the WDH conducted outreach to hospitals and clinics around the state to increase patient referrals to the program.

HCIA also funded capitated payment to pharmacists for each Virtual Pharmacy program participant, and the program coordinator led two days of training for pharmacists and provided weekly check-in calls. Training covered medication management therapy and motivational interviewing.

2. Implementation effectiveness

In this section, we examine the evidence on implementation effectiveness. We assess implementation effectiveness based on program enrollment, selected service- and staff-related measures, and timeliness, relying on interviews with program administrators and self-reported information included in WIPH's quarterly self-monitoring and measurement reports.

a. Program enrollment

As of March 2015, WIPH served 5,011 participants, 84 percent of its three-year target. Most were participants from WyRCT, although this number also included participants from MDP and the Virtual Pharmacy program. Because PCMH and telehealth patients were not considered direct program participants, they are excluded from this enrollment tally. We describe enrollment targets and participation by program below. Unless otherwise noted, figures are based on self-reported data from March 2015.

- The PCMH program recruited 20 primary care practices, double its initial target of 10 practices. These 20 practices serve approximately 130,000 patients. By February 2014, the program had lost two clinic participants that decided not to apply for NCQA PCMH recognition, although it gained two additional clinics, bringing the total to 20 transforming practices.

- WyRCT and TACT nurses assisted 4,292 at-risk participants at 14 hospitals dispersed throughout Wyoming, covering 14 cities within 13 counties and one hospital in Nebraska. One hospital discontinued the program in 2014, after enrolling two participants, bringing the total number of participating hospitals to 13.
- The MDP program dispensed free prescriptions to approximately 500 eligible participants throughout Wyoming. WIPH reported dispensing medications by mail to 374 participants; the rest received medication at dispensing sites.
- The awardee recruited hospitals and primary care clinics into the telehealth program via the Wyoming Hospital Association, outreach, and word of mouth. The program installed telehealth equipment around the state, across 32 communities, spanning all 23 counties and one Nebraska county. Although the awardee did not provide a tally of patients accessing telehealth, WIPH reported 426 telehealth-related health care claims for Cheyenne Regional Medical Center.
- The awardee recruited 12 pharmacies to the Virtual Pharmacy program. Exact numbers are unavailable, but WIPH reported low patient enrollment when the program was discontinued in July 2014.

b. Service measures

WIPH did not specify targets or service-related goals for all program components; however, it did track and monitor trends in service delivery and utilization.

As of June 2015, ten PCMH sites obtained NCQA PCMH recognition and participating PCMH sites collectively increased use of four of seven tracked preventive services. WIPH reported modest increases for tobacco use assessments, tobacco cessation prevention, colorectal cancer screening, and cervical cancer screening from June 2013 to December 2014. During the same period, the awardee reported reduced childhood immunization and ADHD prescription follow-up rates, and no changes in breast cancer screening rates (Table II.A.3). Although practice participation in the HCIA program might have contributed to the reported increases (or decreases) in preventive service use, many other factors could have contributed as well—or, indeed, been the sole reasons for the changes—such as a practice's own motivation (independent of participating in the program) to improve clinical performance, or beneficiaries becoming more aware of the lack of copayments for many Medicare-covered preventive services after the Affordable Care Act.

The WyRCT program met its service goal for medication reconciliation. WyRCT reported that it adhered to the strict intervention protocol and completed medication reconciliation for 100 percent of participants served from October 2013 to March of 2015.

Telehealth assessed program success by the number of installations and by use of the technology among clinical areas. As of March 2015, WIPH reported 179 webcams in hospitals and 239 in physician offices, 2,914 hours of training use, and 6,966 hours of provider-to-provider tele-consultations. To increase clinical applications of telehealth, WIPH encouraged clinics and hospitals to use trainings and administrative meetings to familiarize staff with the technology.

Table II.A.3. PCMH process improvement metrics

Measure	June 2013	December 2014	Percentage point change
Percentage of participants who smoked who received tobacco cessation intervention (NQF 0028)	26%	37%	11
Percentage of participants given a tobacco use assessment (NQF 0028)	72%	76%	4
Percentage of participants ages 50 to 75 who received colorectal screening (NQF 0034)	22%	25%	2
Percentage of women ages 21 to 64 who received one or more Pap tests to screen for cervical cancer (NQF 0031)	31%	33%	2
Percentage of women ages 40 to 69 who received a mammogram to screen for breast cancer (NQF 0031)	27%	27%	0
Percentage of children age 2 who received recommended vaccinations (NQF 0038)	51%	44%	-7
Percentage of children newly prescribed ADHD medication who had at least 3 follow-up care visits within a 10-month period, one of which was within 30 days of when medication was dispensed (NQF 0108)[a]	63%	52%	-11

Source: Self-reported awardee measurement and monitoring results through December 2014, the last quarter during which the awardee reported these measures.

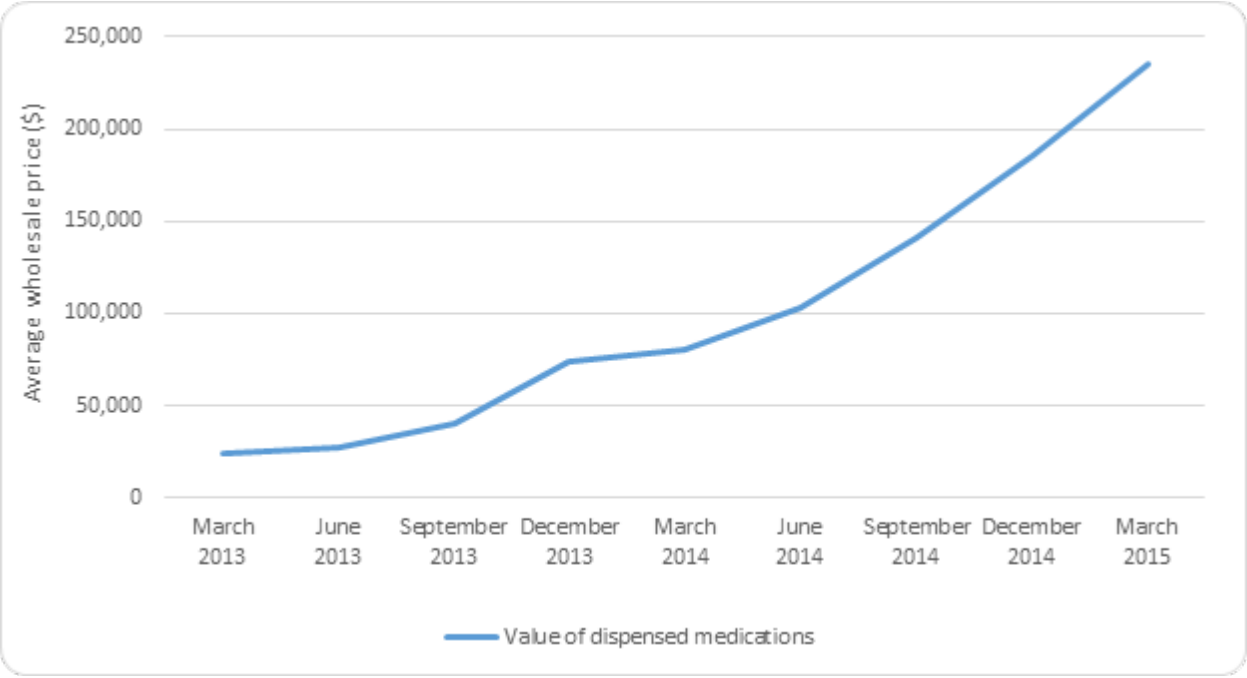
Note: Reported changes from June 2013 to December 2014 are a percentage point difference.

[a] The NQF measures includes two rates: the percentage of children who received follow-up within 30 days and those who received additional follow-up after 30 days. The awardee provided only the initial follow-up rate.

ADHD = attention deficit hyperactivity disorder, NQF = National Quality Forum

The MDP received referrals from 15 hospitals, which fell short of its intended target of 27. However, WIPH reported overall growth for the program, including a 10-fold increase in the value of medications dispensed, from $23,947 in the first quarter of 2013 to $235,658 in the first quarter of 2015 (Figure II.A.1).

Figure II.A.1. MDP value of medications dispensed



Source: Awardee measurement and monitoring results reported by WIPH through March 2015.
Note: Evaluators did not verify data reported.

c. Staffing measures

Awardee documents suggest that as of March 2015, WIPH did not meet its training target of 93 staff, and ultimately trained 60 to 63 HCIA-funded employees. WIPH trained 51 WyRCT nurse coaches, and 9 to 12 pharmacists. In addition, the awardee trained about 45 non-HCIA-funded staff. The awardee did not provide additional information about the type of non-HCIA-funded staff trained or the content of these trainings.

As of March 2015, the WyRCT and TACT programs employed 26 nurses. Some hospitals reported challenges recruiting and retaining WyRCT nurses (discussed in more detail in Section II.3.c), which accounts for the difference between the number of nurse coaches trained and deployed (Table II.A.4).

Table II.A.4. WIPH staff trained and deployed as of March 2015

Staff	Number deployed	Number trained	Duration of training (hours)	Total hours of training
WyRCT nurse coach	26	51	16	816
Pharmacist[a]	0	9–12	16	108–144
Other[b]	--	33–36	16	528–576

Source: Awardee measurement and monitoring results reported by WIPH through March 2015.

Note: Evaluators did not verify data reported.

[a] The awardee reported in its June 2014 narrative that it had met its milestone of training 12 pharmacists. In contrast, quarterly reports through June 2014 suggest that 9 pharmacists were trained.

[b] A note in the December 2015 quarterly report suggests that earlier reports of the number of trainees did not include non-HCIA-funded staff. The awardee does not specify the type of staff, relevant program component, or content of trainings.

d. Program time line

WIPH implemented most program components on schedule, although NCQA PCMH recognition and virtual pharmacy implementation were both delayed. The target date for PCMHs to submit their applications for NCQA recognition was January 2015. During April 2014 site visits, physicians at transforming practices expressed doubt that all participating clinics—particularly smaller, independent practices—could meet the requirements of the NCQA application by the end of the award. Indeed, two clinics notified WIPH in July 2014 that they would not move forward with the NCQA application. One clinic cited the facility's transition to a new electronic health record (EHR) and the other cited a lack of capacity to apply as reasons for withdrawal. During our site visits in April 2015, clinic documentation and discussions revealed that overall, 13 practices received or were in the process of applying for NCQA recognition, and the remaining 7 practices decided not to pursue NCQA recognition during the award period (Table II.A.5).

Table II.A.5. NCQA PCMH recognition status for participating practices

NCQA PCMH recognition status	Number as of April 2015
Achieved level 3 NCQA recognition	3
Submitted NCQA applications, awaiting decision	6
Preparing application	4
Will not pursue	5
Might pursue after HCIA award period	2

Source: Interviews from second site visit, April 2015. In June 2015, awardee documents indicated that 10 practices had received NCQA recognition. Documents did not specify the level of recognition and status of remaining sites.

Note: Evaluators did not verify data reported.

The awardee experienced its most significant delays with implementation of the Virtual Pharmacy program. Program administrators cited early delays related to an extended vacancy in

the team manager position and administrative hurdles associated with UW's internal hiring and contracting processes. Program leadership also cited challenges related to recruiting pharmacies and connecting them to PCMHs, such as delays associated with executing legal agreements, reconfiguring pharmacists' workspaces, and information sharing between pharmacists and PCMHs. The Virtual Pharmacy program targeted non-dually eligible, adult Medicaid patients with certain chronic conditions, which restricted the pool of eligible patients. Furthermore, pharmacists expressed difficulty in recruiting and engaging the eligible population, citing patients' challenges such as transportation issues and intermittent telephone service; as a result, enrollment in the program remained low. CMMI discontinued funding for the Virtual Pharmacy program and the program ended in July 2014.

3. Implementation experience

In this section, we review four domains associated with implementation experience: (1) program characteristics, (2) implementation process, (3) internal factors, and (4) external environment. Implementation research has shown that barriers and facilitators within these domains are important determinants of implementation effectiveness. Table II.A.6 summarizes the major facilitators and barriers to WIPH's implementation effectiveness in each domain. This section focuses exclusively on PCMH and WyRCT, the two largest components of the award.

a. Program characteristics

Two characteristics of WIPH's initiative influenced program implementation: (1) a shared perception that the PCMH model of care represented a relative advantage over traditional care delivery models and (2) the WyRCT program design.

First, providers across sites reported that the PCMH model of care represented an improvement over traditional approaches, facilitating program implementation. As one practice administrator described during our April 2015 site visit, "It's the right thing to do. Providers need to help patients to get better." Participating staff identified three program characteristics as key facilitators of implementation; compared to the prior approach (1) participants were more satisfied with the availability of same-day and evening and weekend appointments, (2) staff were more satisfied with team huddles and previsit planning, and (3) providers appreciated information obtained from new reports, such as the percentage of patients due for colorectal exams or trends in no-show rates.

Second, in the WyRCT program, nurses and administrators viewed the home visit requirement as a critical implementation facilitator. Nurses described participants in the hospital as overwhelmed and often medicated, often resulting in post-discharge confusion about medications and care plans. WyRCT nurses reported that home visits enabled them to identify risks in participants' homes and to provide successful coaching and mentoring for them. Nurses also reported that through home visits, they could conduct a more complete review of participants' medications, enabling more comprehensive medication reconciliation and counseling. Stakeholders we interviewed believed the home visit improved the effectiveness of care transition services provided to participants.

Table II.A.6. Facilitators and barriers to implementation effectiveness

	PCMH		WyRCT	
Domain	Facilitators	Barriers	Facilitators	Barriers
Program characteristics	Perceived relative advantage of PCMH approach to care, including: • Availability of same-day and evening/weekend appointments • Team huddles and previsit planning • Patient reports	• None identified	• Program requirement that nurses complete home visits	Program elements, including: • Qualifying condition and age-based enrollment restrictions • Service area restrictions for home visits • Requirements for nurses to be available at the hospital 7 days per week, and to complete home visits within 48 hours of a particiapant's discharge • Voluntary patient enrollment
Implementation process	Increased physician engagement via: • TransforMED's learning collaboratives and assistance with NCQA applications • Reports on quality measures • Reduced workloads	• Lack of physician engagement/ physician champion • Distribution of HCIA funds	• Physician engagement via on-site nurse presence and word of mouth	• Lack of physician engagement/familiarity with the program
Internal factors	• Technical aptitude	• Lack of technical aptitude • Lack of staff capacity	• Leadership of program coordinator	• Lack of staff capacity
External factors	• National health policy trends toward value-based care	• Insufficient reimbursement for new care processes	• None identified	• Insufficient local resources to meet participants' needs

Source: Interviews from second site visit, April 2015; document review, March 2015.

Program administrators and frontline staff identified other aspects of the WyRCT program design that acted as barriers to implementation. For example, some stakeholders suggested the target population was too narrowly defined, believing that some at-risk patients who did not fit the age and qualifying conditions requirements would have benefited from the program. The WyRCT program generally recommended that hospitals restrict eligibility to patients living within 50 miles of the discharging hospital, although hospitals could specify their own service area criteria. Some nurses identified the service area restriction for home visits as a barrier,

noting that patients who live farther from the hospital, in more remote and rural areas, were at higher risk. There were also staffing challenges that served as barriers to implementation. These included requirements for nurses to be available at the hospital seven days per week, and to complete home visits within 48 hours of a participant's discharge. Rural hospitals struggled to hire WyRCT nurses, especially hospitals with one or fewer FTE staff allocated to the WyRCT program (discussed in greater detail in Section II.3.c). Some nurses also cited voluntary patient enrollment as problematic, suggesting that physician referral to the program would lead to higher enrollment. Finally, interviews suggested that coordination between primary care physicians and health coaches could have been better. The WyRCT protocol directed nurses to coordinate with participants' PCMHs but did not provide specific instructions for doing so. Nurse coaches accompanied some participants to primary care appointments, and reported that their presence confused some physicians. In some cases, WyRCT participants' primary care physicians were also participating in the PCMH intervention. PCMH physicians with whom we spoke reported minimal familiarity with WyRCT, and two expressed a desire for more consistent coordination with nurse coaches. WIPH piloted TACT in response to many of these challenges. The program is located in outpatient facilities, and physicians refer participants to the program. TACT also added depression and anxiety to the list of qualifying conditions and made the program available to patients ages 18 and older.

b. Implementation process

Two primary implementation process factors affected the implementation of WIPH's medical neighborhoods program: (1) engaging physicians at both transforming PCMHs and WyRCT hospitals and (2) dedicating resources to PCMH transformation. First, engaging physicians at PCMHs to adopt the new model of care presented some early challenges, especially for clinics that lacked a designated physician champion to advocate and lead the transformation. However, program staff cited aspects of the implementation process that facilitated physicians' engagement. For example, providers expressed appreciation for TransforMED's learning collaboratives, which offered an opportunity to learn from other sites. Awardee staff also described providers as more engaged as they moved through implementation and began to see benefits of new care processes, such as improved scores on quality measures. Some physicians noticed reduced workloads after shifting some tasks, such as participant follow-up, to nurses.

Engaging physicians at WyRCT hospitals also presented some challenges that staff worked to overcome. Initially, physicians at hospitals and outpatient facilities lacked familiarity with the program. WyRCT nurses cited their on-site program presence and word of mouth as factors related to increasing advocacy for the program and patients' enrollment. Although outpatient physicians at PCMH sites indicated they were not as involved with the WyRCT program, anecdotes from TACT nurses suggested that, at the two outpatient clinics piloting TACT, physician buy-in and patient referrals increased over time. Nurses attributed this uptake to physicians observing reduced visits from frequent users following nurse interventions.

Second, the distribution of HCIA funds also presented a challenge for the PCMH program. The awardee allocated HCIA resources to TransforMED's practice facilitation services, telehealth equipment at participating clinics, and small grants to help pay NCQA application fees. However, the awardee did not provide HCIA funding to participating practices for staff or

EHR upgrades. As the program progressed, staff at participating practices indicated that TransforMED practice facilitation services improved when the focus of the service shifted to the NCQA application submission process. Clinic staff cited TransforMED's application document review, via an electronic upload tool, as the most helpful feature of TransforMED's services, and believed it facilitated implementation of the PCMH model. Several participating PCMH staff commented that funding for EHR-related tasks would have further facilitated implementation.

c. Internal factors

Characteristics of the organization implementing a program can influence implementation effectiveness. Three internal factors affected implementation of the PCMH and WyRCT programs: (1) staff competencies with technology and EHR systems, (2) clinic staffing capacity and workload management, and (3) program coordinator leadership.

First, WIPH leadership stressed the importance of hiring or allocating staff with high technical aptitude or competencies with EHRs. Many sites said the PCMH transformation, which required creating new types of patient reports and new EHR processes, was administratively burdensome, but that the availability of qualified staff helped overcome this implementation challenge. One independent, level 3 NCQA PCMH practice said that a dedicated care coordinator staff position was vital to the success of its transformation. A local foundation sponsored the care coordinator position and the care coordinator led the EHR-based work.

Conversely, although WIPH leadership viewed technical aptitude as a critical facilitator, the lack of technical aptitude was a substantial PCMH implementation barrier, especially for clinics transitioning to new EHRs. For example, clinics without dedicated staff said that extracting data from EHRs was problematic. Several stakeholders we interviewed cited specific challenges within their clinics. One clinician believed that the EHR lacked the functionality necessary to support the program. Many practices reported difficulties funding an EHR-dedicated position and incorporating transformation and EHR tasks into existing billing and reimbursement procedures. Another clinic transitioned to a new system early in the award, after learning its existing EHR would not be certified for Meaningful Use, and cited the EHR as the biggest challenge to PCMH implementation. At least two clinics decided not to pursue NCQA recognition during the award after transitioning to a new EHR.

Second, clinic staff capacity and workload management presented a related challenge for PCMH implementation. PCMH staff across sites agreed that the transformation process was time-consuming and it was difficult to find staff with available time and flexibility to lead and implement change. Smaller clinics with fewer nurses per physician reported challenges implementing previsit planning, same-day appointments, and patients' care plans. One hospital-based clinic hired new staff to address this challenge, but speculated that independent clinics were less likely to have sufficient resources to hire. Respondents recommended that clinics have adequate staff in place before transforming into a PCMH.

Inadequate staff capacity also created challenges for WyRCT. Hospitals reported difficulty providing daily nurse coverage and follow-up within 48 hours of hospital discharge. Although WIPH reported that nurse-to-patient ratios were reasonable based on discharge volume, hospitals

with fewer FTE nurses reported difficulty distributing nurses' hours to ensure daily coverage. Smaller hospitals also considered the seven-day coverage requirement a challenge in staff recruitment. Nurse applicants did not want to work seven days a week to be responsible for the coverage on their own, nor did they want to work less than full time as a way to provide the coverage as part of a small care team. In response to staffing challenges, some hospitals trained nurses who were not HCIA-funded to provide back-up support.

Third, WyRCT and TACT nurses cited the importance of the program coordinator to successful implementation. The program coordinator led trainings, provided ad hoc support, and compiled participants' data. The program coordinator's leadership ensured consistency in the intervention across several disparate settings.

d. External factors

Features of an organization's external environment can also influence program implementation. Two primary external factors affected PCMH and WyRCT implementation: (1) enhanced payments for PCMH practices and (2) WyRCT participants' complex needs amidst scarce community resources.

First, several programs offer incentives to practices that adopt PCMH principles. NCQA recognition provides opportunities for enhanced Medicaid payment in Wyoming; Medicare offers a chronic care management code; and one private insurer, Blue Cross Blue Shield of Wyoming, offers enhanced payment to PCMHs that agree to use the payer's data management tool. Notably, leadership at Cheyenne Regional Medical Center said they anticipate incorporating Medicare's chronic care management payments into the upcoming year's operational budget.

However, practices also commented on challenges associated with taking advantage of these payment opportunities. One practice, which achieved NCQA recognition, reported that reimbursement rewards were insufficient to offset the practice's additional costs, and believed smaller practices were disadvantaged when negotiating with payers. This clinic attempted to negotiate with a private insurer for enhanced payment, but the clinic did not qualify due to small patient panel size—a common characteristic among rural clinics. The same clinic had limited EHR interoperability in the region and did not meet the technology requirement to qualify for Medicare's new chronic care management code. This practice also cited misaligned clinical reimbursement policies as a related challenge. For example, some clinical care practice guidelines advise physicians to test the hemoglobin A1c levels of patients' with diabetes twice per year (American Diabetes Association 2015). However, because insurance does not always cover the second test, many patients are reluctant to schedule visits for which they are responsible for the out-of-pocket payment. Despite challenges, providers predicted that practicing patient-centered care was harmonious with national health policy trends toward value-based reimbursement.

Second, participants' complex needs and the scarcity of community resources presented challenges to WyRCT nurses. The WyRCT program targeted a high-risk population in rural communities. Especially in remote parts of Wyoming, community-based resources such as senior

housing and transportation services were often unavailable. WyRCT nurses traveled extensively for home visits, covering large geographic areas, often up to 50 miles from the hospital that discharged the participant. The home visit, though cited as essential, placed additional demands on nurses' schedules and compounded the problem of inadequate staffing for some sites. Nurses reported that many WyRCT participants did not have the means to purchase necessary medications or equipment, or lacked the motivation to effectively manage their chronic condition(s), despite coaching. WyRCT nurses were also challenged to avoid duplication of efforts for participants who received home health care or resided in assisted living facilities.

4. Sustainability and scalability

Four of the programs will continue in some capacity after the award: (1) PCMH, (2) WyRCT, (3) telehealth, and (4) MDP. (As discussed earlier, the Virtual Pharmacy program was discontinued in July, 2014.)

The PCMH model of care delivery facilitated new processes in participating clinics, such as increased patient access and new population management reports. For practices that achieved or applied for NCQA recognition, the patient-centered approach will continue, largely due to practices' internal motivation to provide quality care. WIPH expects that as national policies trend toward value-based payment models and as staff become savvier with their EHRs, the PCMH approach to care will expand in the state. However, clinic staff expressed concern the model might not be feasible for small, independent practices unless payment policies to support nonbillable aspects of the PCMH are put in place.

Despite the operational challenges, nine of the hospitals that implemented the WyRCT program plan to continue to employ nurses to provide transitional care services. To ease program implementation, many plan to change the original eligibility criteria and staffing requirements of the program. WIPH framed its argument for sustaining WyRCT around new chronic care management billing codes and cost avoidance. For example, administrators at Cheyenne Regional Medical Center were exploring whether the hospital could demonstrate cost-avoidance associated with reduced hospitalizations for uninsured and Medicare recipients receiving care transition services. Cheyenne Regional Medical Center also plans to continue the TACT program.

WIPH foresees that hospitals and clinics will continue using telehealth equipment, especially at large medical centers where the cost is absorbed. Smaller clinics might try to charge a fee to sustain telehealth. Medicaid and Blue Cross Blue Shield of Wyoming also reimburse clinics for some telehealth encounters, supporting sustainability.

The MDP will continue post-award with continued funding from the WDH.

B. Clinicians' attitudes and behaviors

1. HCIA Primary Care Redesign Clinician Survey

Information gathered from interviews with program leadership and frontline staff at selected clinical sites or satellite offices provided important insights into the implementation process.

Although these in-person interviews provide a rich source of data, views from the leadership and staff are limited to a small number of clinical locations and might not reflect the perspectives of clinicians practicing at other sites. To assess perspectives of clinicians more broadly, we administered the HCIA Primary Care Redesign Clinician Survey to clinicians in the fall of 2014, the third year of the HCIA-funded program. Data from the survey provide additional insights into the implementation process and experience, as well as the contextual factors that might affect implementation effectiveness.

In this section, we report on participating PCMH clinicians' views of their daily work life and practice. Clinicians associated with the other four program components were not surveyed. First, we focus on the contextual factors that can affect program implementation, including the characteristics of the practice location, career satisfaction and burnout, and barriers to providing high quality and patient-centered care, as well as clinicians' perceptions of how well the care team functions. We then present data on the alignment of PCMH clinicians' views and experiences with the overall goals of the HCIA-funded innovation, as well as their awareness of and participation in the Wyoming Medical Neighborhoods program and their views of the barriers and facilitators of successful program implementation.

2. Contextual factors that can affect successful implementation of the HCIA program

a. Characteristics of clinicians' practice location

A total of 83 PCMH clinicians responded to the survey, resulting in a response rate of 80 percent. The number of clinicians in each response category does not always sum to 83, here and throughout this section, due to survey item nonresponse, as well as clinicians who reported that a given question did not apply to their practice and thus did not provide a response. In addition, data is not included in the tables for survey responses with fewer than 11 respondents. Table II.B.1 shows the distribution of respondents by types of clinician, types of practice, and primary compensation source.

Participating Wyoming clinicians reported working in settings that were above average in terms of health information technology. Nationally, slightly more than half of physicians practice in settings with functional EHRs (Furukawa et al. 2014), but most participating Wyoming clinicians surveyed reported using health information technology at their practice locations. Table II.B.2 shows that most Wyoming clinicians used EHR systems for various functionalities, including use of electronic tracking systems and patient registries, advanced functions that are not in widespread use nationally (DesRoches, Painter, and Jha 2014). Wyoming also offers patient-facing technologies such as electronic prescription refills and appointment requests. The application process for NCQA PCMH recognition, which requires practices to adopt many of these electronic functionalities, might have facilitated this higher than average use of health information technology.

Table II.B.1. Types of clinicians, practices, and compensation sources

Survey item	Number of respondents	Percentage of respondents
Type of clinician		
Physician	44	54%
Nurse practitioner	12	15%
Physician assistant	24	29%
Type of practice		
Group practice (3 or more clinicians)	46	55%
Federally Qualified or other community health center	21	25%
Other (hospital-based practice, solo practice, other)	13	16%
Primary compensation source		
Fixed salary	41	49%
Salary adjusted for performance	19	23%
Other (hourly/time-based, fee-for-service, other)	20	24%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Table II.B.2. Electronic capabilities for clinicians and patients

Survey item	Number of respondents	Percentage of respondents
Physicians using EHR to		
Access laboratory results	80	96%
Receive drug dosing and interaction alerts	80	96%
Enter clinical notes	80	96%
Order tests and procedures	79	95%
Prescribe medications	77	92%
Track electronic referrals	66	80%
Access participant registries	60	72%
Participants can		
Email clinician about a medical question or concern	53	64%
Refill prescriptions	44	53%
Request appointments	38	46%

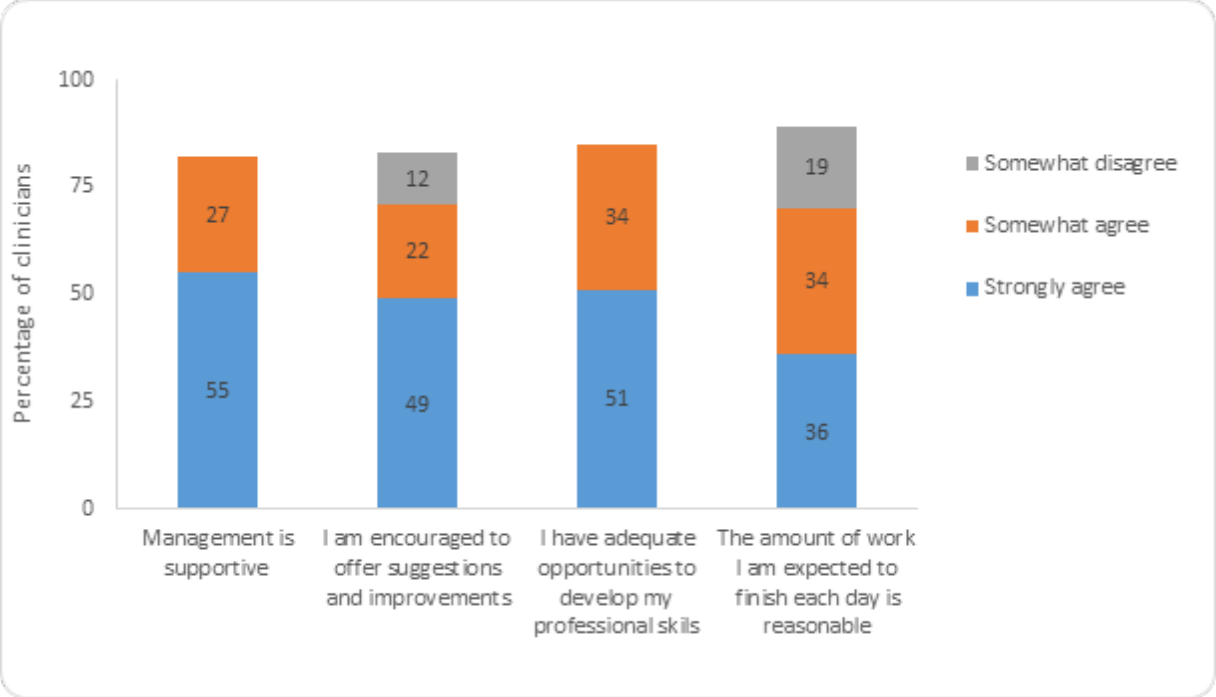
Source: HCIA Primary Care Redesign Clinician Survey, 2014.

b. How clinicians experience their careers and workdays

Clinicians' satisfaction with their overall careers, level of burnout, and perceptions of their practice environments can all have an effect on the success of program implementation and organizational change. Most Wyoming clinicians were at least somewhat satisfied with their overall careers (79 percent) and reported having no symptoms of burnout (55 percent). However, among those who reported that they were satisfied, only 36 percent reported being very satisfied and 40 percent of all respondents experienced one or more symptoms of burnout at the time the survey was taken, including symptoms that would not go away and feelings of complete burnout.

Wyoming clinicians were also generally satisfied with their workplace management. About half of them strongly agreed that their management team was supportive, that they were encouraged to offer suggestions and improvement, and that they had adequate opportunities for professional development. About a third strongly agreed that the amount of work they were expected to complete each day was reasonable (Figure II.B.1).

Figure II.B.1. Workplace ratings



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Totals may not add to 100% due to survey item non-response.

In addition to workplace ratings, the survey included items that assessed clinicians' beliefs about their ability to provide high quality care. Among responding clinicians, 60 percent agreed with the statement, "It is possible to provide high quality care to all of my patients," whereas only 20 percent agreed with the statement, "I lack adequate information from research evidence to guide my clinical decisions." Wyoming clinicians also reported barriers to providing optimal,

patient-centered care. In order of importance, the greatest limitations to the ability to provide timely, patient-centered care were (1) lack of adequate reimbursement, (2) lack of timely information about patients cared for by other physicians, (3) not having enough time with patients during visits, (4) patients' difficulty paying for needed care, and (5) receiving too many reminders from the EHR (Table II.B.3). The overwhelming perception of inadequate reimbursement is notable because most respondents received a fixed or enhanced salary; only 2 percent received fee-for-service reimbursement. In-person discussions with Wyoming frontline staff echoed this finding; clinicians expressed frustration with the amount of funding they received from the Wyoming Medical Neighborhoods program.

Table II.B.3. Perceptions of ability to provide high quality care

Survey item	Number of respondents	Percentage of respondents
It is possible to provide high quality care to all of my patients		
Strongly agree	15	18%
Somewhat agree	35	42%
Neither agree nor disagree	--	--
Somewhat disagree	12	14%
Strongly disagree	--	--
Percentage reporting each of the following at least somewhat limits their ability to provide optimal patient-centered care		
The level of reimbursement is not adequate	75	90%
I do not have enough time to spend with patients during visits	68	82%
I lack timely information about the patients I see who have been cared for by other physicians	61	73%
My patients have difficulty paying for needed care	56	67%
I receive too many reminders from my EHR	49	59%
It is difficult for me to obtain specialized diagnostic tests or treatments for my patients in a timely manner	37	45%
It is difficult for me to obtain specialist referrals for my patients in a timely manner	36	43%
I lack adequate information from research evidence to guide my clinical decisions	17	20%

Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are suppressed because of confidentiality restrictions.

c. Clinicians' perceptions of care team functioning

Most Wyoming clinicians (80 percent) reported working as part of a care team and, overall, their perceptions of how these teams functioned was positive. Among those clinicians working in a team, a majority agreed that members of the care team relayed information in a timely manner (92 percent), used common terminology when communicating with one another (88 percent), had sufficient time for patients to ask questions (83 percent), verbally verified information they

received from one another (75 percent), and followed a standardized method of sharing information when handing off patients (61 percent).

d. Alignment with goals of PCR

The survey included several items asking clinicians to rate the importance of a series of goals related to PCR on a scale ranging from extremely important to not important at all. The inclusion of the extremely category rating helps to provide variation in the data, forcing respondents to choose between goals that are essential to meet and those that are merely important. Wyoming clinicians' responses somewhat align with the goals of PCR (Table II.B.4). Most clinicians rated 4 of the 13 goals as extremely important: (1) increasing patients' capacity to manage their own care, (2) reducing ED visits, (3) increasing access to primary care, and (4) improving care continuity in primary care. Notably, very few Wyoming clinicians rated "increasing the number of primary care practices functioning as a patient-centered medical home" as extremely important, although a specific goal of the Wyoming Medical Neighborhoods program is to create PCMHs. This result indicated a perception that the PCMH designation was less of a priority for clinicians than improving care for their patients.

3. Awareness of program, receipt of training, and perceived effects

The overall goal of the Wyoming Medical Neighborhoods program was to change the way care was provided by embedding PCMHs into medical neighborhoods. Program administrators believed that clinicians were critical to that process, particularly physician champions who could lead reform within practices. Because the awardee leveraged partnerships with independent practices across the state, clinician buy-in was critical to successful transformation. Understanding clinicians' perceptions of the program could be a key factor in understanding the effect of the program on patients' outcomes. For example, if clinicians are aware of the program, have received appropriate and effective training, and believe that adopting a medical home model will have a positive effect on the care they provide, they are likely to feel more invested in the program's success. Alternatively, those who feel more negatively about the program, often due to competing priorities and inadequate resources, might be less likely to enthusiastically implement the intervention. In this section, we report on Wyoming clinicians' experiences with and perceptions of the Wyoming Medical Neighborhoods program.

a. Awareness of the program and receipt of training

Most (77 percent) of the Wyoming clinicians we surveyed were at least somewhat familiar with the Wyoming Medical Neighborhoods program. Of these clinicians, 66 percent had received training related to the program. On average, clinicians reported receiving 18 hours of program-related training. Trainings included site visits and learning collaboratives with TransforMED. (Site visit interviews suggested some physicians also might have included internal, practice-led trainings in their responses.)

Table II.B.4. Importance of PCR goals

Survey item	Number of respondents	Percentage of respondents
Percentage of clinicians rating each of the following as extremely important:		
Improving patients' capacity to manage their own care	46	55%
Reducing ED visits	44	53%
Increasing access to primary care	44	53%
Improving care continuity in primary care	43	52%
Improving care coordination for patients with chronic conditions	41	50%
Improving appropriateness of care	38	46%
Increasing the use of evidence-based practice in clinical care	35	42%
Reducing hospital readmissions	33	40%
Reducing overall health care spending	31	37%
Improving capability of health care organizations to provide team-based care	29	35%
Increasing use of EHRs and other health IT	26	30%
Improving the capability of health care organizations to provide patient centered care	15	18%
Increasing the number of primary care practices functioning as a PCMH	--	--

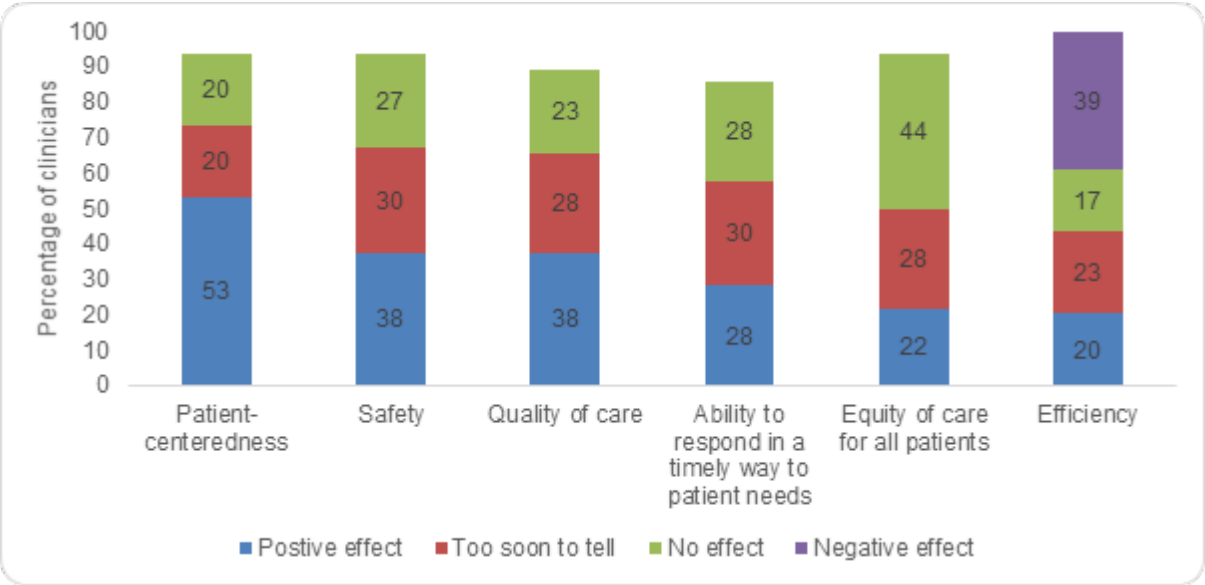
Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Figures are based on item response and those who believed that the question applied to their practice. Survey items with fewer than 11 respondents are suppressed because of confidentiality constraints.

b. Perceived effect of program on patient care

Clinician's perceptions of the effect of the Wyoming Medical Neighborhoods program on the care they provided to patients were mixed. Clinicians were asked about the perceived effect of the Wyoming Medical Neighborhoods program and the barriers to and facilitators of implementation only if they reported being at least somewhat familiar with the program. About half of the clinicians who were familiar with the program believed it would have a positive effect on the patient-centeredness of the care they provided, and nearly a third believed the program would have a positive effect on their ability to respond to patient needs in a timely way. Conversely, more than a third of clinicians felt that the program was having a negative effect on the efficiency of patient care (Figure II.B.2). Based on our in-person discussions with clinicians, concerns about efficiency likely are related to EHR challenges, such as increased documentation requirements, although this is speculative, given that we did not ask about reasons for this belief in the survey. Some clinicians believed that the intervention would have no effect or that it was too soon to discern an effect on patient safety, quality of care, and equity of care, and clinicians' ability to respond in a timely way to patient needs.

Figure II.B.2. Perceptions of effects of program on patient care



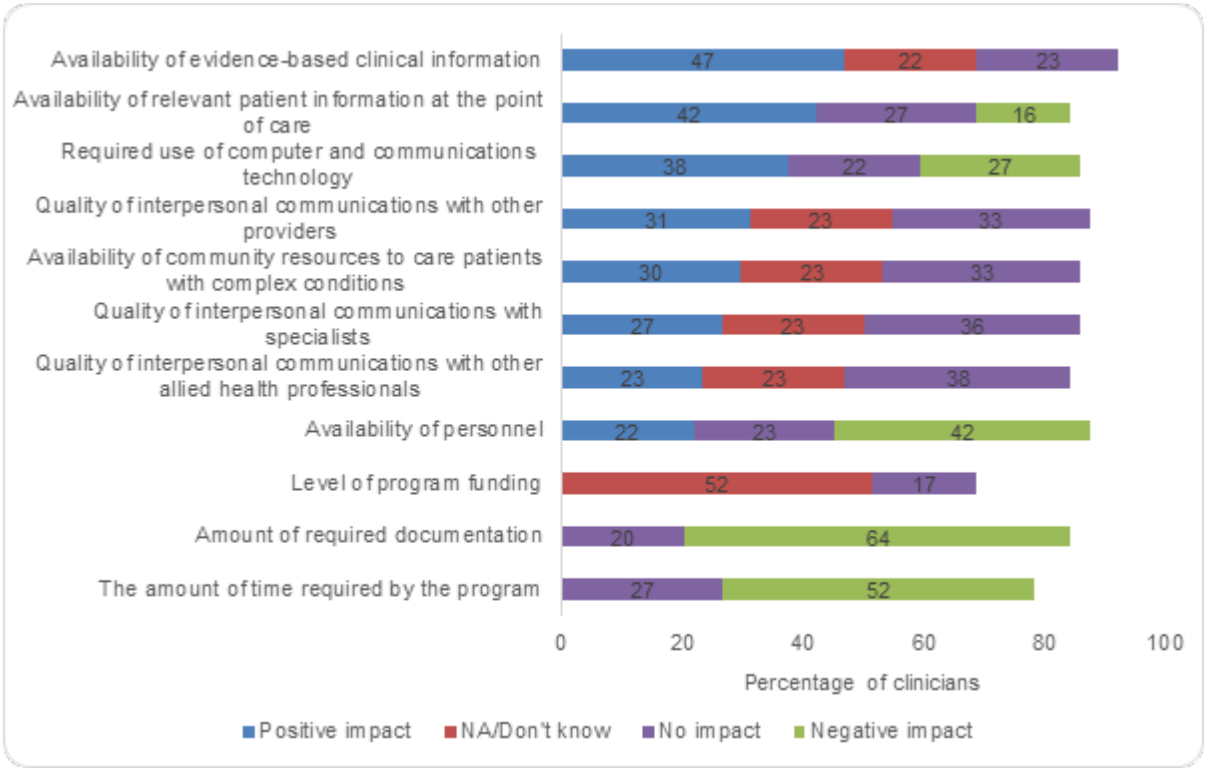
Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Totals may not add to 100% due to survey item non-response.Figures are based on the total number of Wyoming clinicians reporting they were at least somewhat familiar with the Wyoming Medical Neighborhoods program.

c. Barriers to and facilitators of program implementation

Finally, we asked Wyoming clinicians who were at least somewhat familiar with Wyoming Medical Neighborhoods to rate the effect of a series of barriers to and facilitators of program implementation. Clinicians were most likely to cite availability of evidence-based clinical and patient information at the point of care as facilitators to program implementation. Clinicians were ambivalent about the effect of the quality of interpersonal communication with other allied health professionals, specialists, and other providers, with most selecting no impact or not applicable/don't know. The most often-cited barriers to program implementation were the amount of required documentation, the amount of time required by the program, and the availability of necessary personnel (Figure II.B.3). Clinicians' concerns about burdensome documentation requirements are consistent with their views that the program negatively affects efficiency of care.

Figure II.B.3. Barriers to and facilitators of program implementation



Source: HCIA Primary Care Redesign Clinician Survey, 2014.

Note: Survey items with fewer than 11 respondents are not shown because of confidentiality restrictions. Totals may not add to 100% due to survey item non-response. Figures are based on the number of clinicians who reported being at least somewhat familiar with the Wyoming Medical Neighborhoods program.

4. Conclusions about clinicians' attitudes and behavior

Wyoming clinicians surveyed reported higher-than average EHR capabilities to support Wyoming Medical Neigborhoods. The majority of clinicians expressed career satisfaction, although 40 percent were experiencing at least one symptom of burnout. Most physicians felt they could provide patient-centered care, but cited barriers to care such as lack of adequate reimbursement, lack of timely information about patients cared for by other physicians, and not having enough time with patients during visits. A majority of clinicians reported working in well-functioning care teams. Collectively, the three most important PCR goals to clinicians were (1) improving patients' capacity to manage their own care, (2) reducing ED visits, and (3) increasing access to primary care. The most commonly cited faciliatators to program implementation were the availability of evidence-based clinical information and relevant patient information at the point of care. Conversely, a majority of surveyed clinicians felt that the amount of documentation and time required by the program were barriers to implementation.

C. Impacts on patient outcomes

1. Introduction

In this part of the report, we present preliminary results for the WIPH-PCMH HCIA program on patient outcomes in three domains: quality-of-care outcomes, service use, and spending. Although the WIPH-PCMH program serves Medicaid beneficiaries and Medicare beneficiaries enrolled in managed care plans as well as Medicare fee-for-service (FFS) beneficiaries, due to limitations in available data we have analyzed outcomes only for the Medicare FFS population (including those who are dually eligible for Medicare and Medicaid). Results might not be generalizable to the full population that the WIPH-PCMH program serves. As described in Section II.A.2., practices joined the PCMH program at two different times. For the impact evaluation, we define practices as part of one of two cohorts based on when they joined the intervention (January 1, 2013 for cohort one and January 1, 2014 for cohort two). We report preliminary results for the 18 cohort one practices only due to lack of available data for the intervention quarters of primary interest for the 2 cohort two practices. In addition, although the PCMH was designed to affect patients covered by any insurance, we have limited the impact analysis to Medicare fee-for-service (FFS) beneficiaries due to limitations in available data. We first describe the methods for estimating impacts (Section II.C.2) and then the characteristics of the 18 cohort one treatment practices at the start of the intervention (Section II.C.3). We next demonstrate that the 18 treatment practices were similar at the start of the intervention to the 69 practices we selected as a comparison group, which is essential for limiting potential bias in impact estimates (Section II.C.4). Finally, in Section II.C.5, we describe the preliminary quantitative impact estimates, their plausibility given the implementation findings, and our next steps before drawing conclusions. We do not draw conclusions about program effectiveness at this time for several reasons; in particular, the current analyses suggest there might be some imbalance (or unmeasured differences) between cohort one treatment and comparison groups that could affect results and require additional investigation or sensitivity analyses. In addition, the analyses do not yet include the cohort two practices and do not cover the full period over which the intervention is expected to have an effect.

As described in Section I, the PCMH program is one of five components in WIPH's overall effort to promote medical neighborhoods in Wyoming. In future reports, we also plan to assess the impact of WIPH's Rural Care Transition's program on patient outcomes. These two program components—PCMH and WyRCT—are central to WIPH's overall intervention. We will not separately assess the impacts of WIPH's other program components—telehealth, the Medication Donation Program, and the Virtual Pharmacy program—because we either lack identifiers for the providers participating in the program, lack claims for the majority of patients benefiting from the program, or are unable to replicate the enrollment criteria, making it difficult to construct a meaningful comparison group.

2. Methods

a. Overview

We estimated program impacts as the difference in outcomes for patients assigned to 18 treatment practices and outcomes for patients assigned to 69 matched comparison practices,

adjusting for any differences between the groups before the PCMH intervention began. In each domain, we specified one or two primary tests before conducting impact analyses. Each primary test defined the outcomes, population, time period, direction of expected effects for which we hypothesize to see impacts if the program is effective, and thresholds that we consider substantively important. We shared these primary tests with CMMI and WIPH, providing them with an opportunity to comment, and we revised the tests as appropriate. We describe the results of these primary tests in the context of the implementation findings and secondary quantitative tests (robustness and model checks). Based on the preliminary findings, we describe additional analyses that will provide further evidence on the robustness of our estimates.

b. Treatment group definition

We defined the treatment group separately in the baseline period before the practices joined the intervention and the period after they joined (the intervention period). The baseline period for cohort one[1] is January 1, 2012, to December 31, 2012, and the intervention period is from January 1, 2013 to December 31, 2014. In each quarter of the baseline or intervention period, the treatment group consists of Medicare FFS beneficiaries who (1) were assigned to one of the cohort one treatment practices on or before the first day of the quarter (see below for attribution and assignment methods); (2) had observable outcomes for at least one day in the quarter; and (3) lived in Wyoming, Nebraska, or Montana for at least one day of the quarter. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

Attribution and assignment method. We attributed beneficiaries to practices using similar decision rules that CMMI uses for the Comprehensive Primary Care (CPC) Initiative. Specifically, in each baseline and intervention month, we attributed beneficiaries to the primary care practice whose providers (physicians, nurse practitioners, or physician assistants) provided the plurality of primary care services in the past 24 months. When there was a tie, we attributed the beneficiary to the practice he or she visited most recently. WIPH provided identifiers for the treatment practices and the providers who worked in them (and when). We obtained data on providers in other practices from SK&A, an outside health care data vendor that maintains and verifies lists of providers who work in practices throughout the country, and we used the SK&A data to supplement the treatment provider data from WIPH.

In each period (baseline and intervention), we *assigned* the beneficiary to the first treatment practice to which he or she was attributed in the period, and continued to assign the beneficiary to that practice for all quarters in the period. For clarification, at the end of the attribution process, beneficiaries can be *attributed* to more than one practice during the baseline and/or the intervention periods; at the end of the assignment process, a beneficiary is *assigned* to the practice to which he or she was first attributed in the relevant period, either baseline or

[1] Eighteen practices are part of the cohort one PCMH intervention, including one practice with two locations and two separate site identifiers that is considered two practices for the purpose of the impact evaluation. We excluded one practice that dropped out of the intervention shortly after it started and never submitted any identifying information, as well as one practice that did not submit any identifying information, so although it is part of the intervention, it is not included in the impact evaluation.

intervention. This rule ensures that, during the intervention period, beneficiaries did not exit the treatment group solely because the intervention succeeded in reducing their service use (including visits at treatment panels). The definition for the baseline period corresponds to that of the intervention period so that, across the two periods, interpretation of the population changes over time should be comparable.

c. Comparison group definition

The comparison group consists of Medicare FFS beneficiaries assigned to 69 matched comparison practices during the baseline and intervention periods. Because the WIPH-PCMH program operates throughout Wyoming and those practices that chose not to participate could differ systematically from those that do, we selected comparison practices from neighboring Montana. We selected comparison practices that were similar to the treatment practices during the baseline period on observable factors that can influence patient outcomes, especially those factors that WIPH used when recruiting practices for the intervention. This section describes how we constructed the matched comparison group whereas Section II.C.4 shows the balance we achieved between the two groups on the matching variables.

We selected the 69 comparison practices in four steps. First, we used data from SK&A to develop a list of potential comparison practices. We also obtained *CMS Certification Numbers* (CCNs; CMS is the Centers for Medicare & Medicaid Services) from the Integrated Data Repository for Federally Qualified Health Centers (FQHCs) and Rural Health Clinics (RHCs). Second, we developed matching variables, defined at the start of the intervention (January 1, 2013), for all treatment and potential comparison practices (N = 217). These variables included characteristics of the practice (for example, the number of primary care providers [PCPs] in the practice and whether the practice is owned by a hospital or health system); and characteristics of Medicare FFS beneficiaries assigned to the practices (for example, mean Hierarchical Condition Category [HCC] score and utilization in the baseline period). Section II.C.4 shows all matching variables and their data sources. Third, we dropped potential comparison practices that were unlike treatment practices because they had (1) NCQA PCMH recognition in the baseline period or (2) an average of fewer than 25 assigned Medicare FFS beneficiaries in the baseline period. We also dropped potential comparison practices that were not appropriate matches for our treatment practices, such as Indian Health Services practices. This resulted in a pool of 204 potential comparison practices.

In the final step, we used propensity score methods to select comparison practices (from the pool of 204) that were similar to the 18 treatment practices on the matching variables. The propensity score for a given practice is the predicted probability, based on all of a practice's matching variables, that the practice is in the treatment group (Stuart 2010). The score collapses all of the matching variables into a single number for each practice that can be used to assess how similar practices are to one another. We matched each treatment practice to one or more comparison practices with a similar propensity score, with the aim of generating a comparison group that is similar, on average, to the treatment group on the matching variables (see Section II.C.4 to assess balance between treatment and comparison groups after matching). We specified that comparison practices had to match exactly to the treatment practices on two characteristics: whether the practice was a health center (including FQHCs and RHCs) and, for the health

centers, whether the practice is participating in the CMS FQHC demonstration program because one of the treatment practices is participating in this demonstration program.

We did not match the nonhealth center practices on one key variable used in other awardee analyses—number of assigned beneficiaries. After consultation with CMMI, we chose not to use this characteristic for matching for the nonhealth centers because we did not have comparable data for the treatment and potential comparison practices on the providers working in practices. To determine the providers working in treatment practices, we used National Provider Identifier (NPI) data from WIPH and SK&A. However, for the comparison practices, we had only SK&A data to determine the providers working in practices. We know that SK&A data do not contain an exhaustive list of NPIs, and missingness could be exacerbated in rural areas. Consequently, we might be underidentifying providers in the potential comparison versus treatment groups, which will lead to underassignment of patients to practices. By requiring balance on the *measured* number of assigned beneficiaries, we could be forcing matches that are, in fact, not similar in patient panel size. Therefore, we decided to use the count of providers from SK&A data for both treatment and potential comparison practices for matching non-health centers, and we did not match on the number of attributed patients. Although we are concerned that SK&A might be undercounting providers, that undercount should be similar for both treatment and comparison practices, making the provider count from SK&A a valid matching variable.

We required each treatment practice to match to at least one, and up to 10, comparison practices and that the ratio of comparison to treatment practices be 3:1. This matching ratio increases the statistical certainty in the impact estimates (relative to 1:1 matching), because it creates a more stable comparison group against which the treatment group's experiences can be compared.

After selecting the matched comparison practices, we then assigned Medicare FFS beneficiaries to these practices in each intervention quarter using the same rules we used for the treatment group and for the comparison group in the baseline quarters. Specifically, we assigned a beneficiary to the comparison group in each intervention quarter if he or she (1) was attributed to one of the 69 matched comparison practices on or before the first day of the quarter; (2) had observable outcomes for at least one day in the quarter; and (3) lived in Wyoming, Nebraska, or Montana for at least one day of the quarter. If a beneficiary was attributed to a treatment and comparison practice (which happened rarely), we assigned him or her to the treatment or comparison group based on the first practice he or she was attributed to in the intervention period.

d. Construction of outcomes and covariates

We used Medicare claims from January 1, 2009, to December 31, 2014, for beneficiaries assigned to the treatment and comparison practices to develop two types of variables: (1) **outcomes,** defined for each person in each baseline or intervention quarter that the person is a member of the treatment or comparison group; and (2) **covariates** that describe a beneficiary's characteristics at the start of the baseline and intervention periods, and are used in the regression models for estimating impacts to adjust for beneficiaries' characteristics at the beginning of the period. We defined two sets of covariates—one at the start of the baseline period and one at the

start of the intervention period—and did not update them for the rest of the period. This avoids the potential bias that could occur if the intervention affected both control variables and outcomes. For example, the intervention might result in greater contact with the health system and earlier diagnoses of diseases and conditions, which could affect both health-related characteristics and outcomes. If we adjust for changes in health-related status during the intervention period, we might adjust away part of the impact of the intervention. Appendix 1 provides details on the methods we used to construct these variables.

Outcomes. We calculated four quarter-specific outcomes that we grouped into three domains:

1. Domain: Quality-of-care outcomes
    a. Inpatient admissions for ambulatory-care sensitive conditions (number/quarter); also called potentially preventable admissions
2. Domain: Service use
    b. All-cause inpatient admissions (number/quarter)
    c. Outpatient ED visit rate (number/quarter); outpatient ED visits are defined as ED visits or observational stays that do not end in a hospital admission
3. Domain: Spending
    d. Total Medicare Part A and B spending ($/month)

Three of these outcomes—all but admissions for ambulatory-care sensitive conditions—are outcomes that CMMI has specified as core for the evaluations of all HCIA programs. The fourth outcome that CMMI has specified as a core outcome, unplanned 30-day hospital readmissions, is not included in our primary tests because WIPH does not explicitly expect to affect readmissions with its PCMH intervention.

Covariates. The covariates, or predictor variables, include (1) whether a beneficiary has each of 18 chronic conditions (including physical health, mental health, and disabilities), created by applying Chronic Condition Warehouse algorithms to claims in the 12 to 36 months (depending on the condition) before the start of the baseline or intervention period; (2) HCC scores; (3) demographics (age, gender, and race or ethnicity); (4) original reason for Medicare entitlement (old age, disability, or end-stage renal disease); and (5) whether the beneficiary is dually enrolled in Medicare and Medicaid.

e. Regression model

We used regression models to implement the difference-in-differences design for estimating impacts. For each outcome, the model estimates the relationship between the outcome and a series of predictor variables, assuming that each of the predictor variables has a linear (additive) relationship with the outcome. The predictor variables include the beneficiary-level covariates (defined in Section II.C.2.d), whether the beneficiary is assigned to a treatment or a comparison practice, an indicator for each practice (which accounts for stable differences among practices in

their outcomes over time), an indicator for each post-intervention quarter, and an interaction of a beneficiary's treatment status with each post-intervention quarter. The estimated relationship between the interaction term and outcomes in the intervention period is the impact estimate. It measures the average difference between outcomes for beneficiaries assigned to the treatment and comparison practices, subtracting out any differences between these groups during the baseline period. Finally, the model quantifies the uncertainty in the impact estimates, allowing for statistical tests that determine whether observed differences in outcomes between the treatment and comparison groups are likely due to chance. The model used robust standard errors to account for clustering of outcomes across quarters for the same beneficiary and a dummy variable for each practice (fixed effects) to implicitly account for clustering of outcomes for beneficiaries assigned to the same practice. Appendix 2 provides details on the regression methods, including descriptions of the weights each beneficiary receives in the model.

f. Primary tests

Table II.C.1 shows the primary tests for WIPH-PCMH, by domain. Each test specifies a population, outcome, time period, expected direction of effect, and threshold that we count as substantively important (expressed as a percentage change from the counterfactual—that is, the outcomes that beneficiaries in the treatment group would have had if they had not received the treatment). The purpose of these primary tests is to focus the evaluation on hypotheses that will provide the most robust evidence about program effectiveness (see Appendix 3 for a detailed description of how we selected each test).

Our rationale for selecting these primary tests is as follows:

- **Outcomes.** WIPH's expected impacts are to reduce ED visits, hospitalizations, and spending (three of the four core outcomes) so our primary tests address these three outcomes. The intervention is also expected to improve quality-of-care outcomes, including reducing hospitalizations for ambulatory care-sensitive conditions so our primary tests also address this outcome.
- **Time period.** WIPH did not specify a time period for intervention impacts. To provide time for the program to be implemented and diffused into practice, we chose to analyze impacts starting one year after the start of the program through the end of the intervention (intervention quarters 5 through 10 (I5 through I10) for cohort one and I5 and I6 for cohort two).
    - **Population.** Because WIPH expects to affect all patients served by the treatment practices, the population for our primary tests includes all Medicare FFS beneficiaries assigned to the treatment practices. We do not plan to include Medicaid beneficiaries, because we anticipate that Medicaid data from Montana, Nebraska, and Wyoming will not be current enough to cover the period for the primary tests.

Table II.C.1. Specification of the primary tests for WIPH—PCMH Program

Domain (number of tests in the domain)[a]	Outcome (units)	Time period for impacts (controlling for baseline differences)[b]	Population	Substantive threshold (impact as percentage of the counterfactual)[c,d]
Quality-of-care outcomes (1)	Inpatient admissions for ambulatory care-sensitive conditions (#/person/quarter)	Average over I5 through I10 for cohort one and I5 and I6 for cohort two	Medicare FFS beneficiaries attributed to treatment practices	-5.00%
Service use (2)	All-cause inpatient admissions (#/person/quarter)			-3.75%
	Outpatient ED visit rate (#/person/quarter)			-5.00%
Spending (1)	Medicare Part A and B FFS spending ($/person/month)			-3.75%

Note: For all primary tests, the expected direction of effect is a decrease relative to the comparison group.

[a] We adjusted the *p*-values from the primary test results for multiple comparisons made within each domain, but not across domains.

[b] The regression models controlled for differences between the treatment and comparison groups during the baseline year when estimating program impacts.

[c] For all-cause hospitalizations and spending, we set the substantive threshold to 75 percent of WIPH's expected effect. For outpatient ED visits, we set the substantive threshold based on evidence from the literature (Peikes et al. 2011). For hospitalizations for ambulatory care-sensitive conditions, for which WIPH did not set an explicit target, we used the reductions in acute care that Peikes et al. (2011) indicated could be feasible among beneficiaries in a patient-centered medical home.

[d] The counterfactual is the outcomes the treatment group would have had in the absence of the HCIA-funded intervention.

ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award; PCMH = patient-centered medical home; WIPH = Wyoming Institute of Population Health.

- **Direction (sign) of the impact estimate.** The primary tests are testing for a reduction, relative to the counterfactual for each of the outcome measures.
- **Substantive thresholds.** Some impact estimates could be large enough to be substantively interesting (to CMMI and other stakeholders) even if they are not statistically significant, and for this reason we have specified thresholds for what we call substantive importance. WIPH expects a 10 percent reduction in the ED visit rate, a 5 percent reduction in the all-cause hospital admission rate, and 5 percent reduction in total spending. For the all-cause hospital admission rate and total spending, the substantive thresholds we chose are 75 percent of WIPH's stated goals and are therefore set to a 3.75 percent reduction. The substantive threshold for the ED visit rate is based on evidence from the literature (Peikes et al. 2011) because it is the smaller of the two threshold options. It is set to a 5 percent reduction in the ED visit rate. Given that WIPH did not explicitly set goals for preventable hospitalizations, our threshold for this outcome is also based on evidence from the literature (Peikes et al. 2011). The threshold is set to a 5 percent reduction in preventable hospitalizations.

Due to limitations in data availability, we were able to conduct the primary tests in this report only partially. Specifically, we estimated impacts only during I5 through I8 for cohort one, since Medicare claims were available only through December 31, 2014. Our third annual report will cover I5 through I10 for cohort one and will include cohort two practices.

g. Secondary tests

We also conducted secondary quantitative tests to help corroborate the findings from the primary tests. This is important because some of the differences observed between the treatment and comparison groups for the primary tests could result from the non-experimental design of our study or random fluctuations in the data. We have greater confidence in the primary results if they are generally consistent with the expected broader pattern of results. Specifically, we estimated the program's impacts on the four outcomes during two additional intervention periods: (1) the first 6 months after the practices joined the intervention (I1 and I2), and (2) months 7 to 12 (I3 and I4). Because we expect few or no impacts in the first few months of the program as practices are implementing the intervention, the following pattern would be highly consistent with an effective program—few to no measured effects in the first two quarters, growing effects in I3 and I4, and the largest impacts in I5 through I8 (the period for the primary tests). In contrast, large differences in outcomes (favorable or unfavorable) in the first year of the program could suggest a limitation in the comparison group, not true program impacts.

h. Synthesizing evidence to draw conclusions

Because results are preliminary and require further exploration, we do not draw any conclusions about program effectiveness of the PCMH program in this report. However, we summarize our preliminary impact analyses as (1) statistically significant favorable effect, (2) substantively important favorable effect, (3) indeterminate effect, and (4) substantively important unfavorable effect. (We cannot conclude that a program has a statistically significant unfavorable effect because, in consultation with CMMI, we decided to use one-sided statistical tests, which

do not test for evidence of program harms.) For reference, Appendix 3 describes whether and how the next annual report will translate these results into program effectiveness conclusions.

3. Characteristics of the treatment group at the start of the intervention

This section describes the characteristics of the treatment group at the start of the intervention (January 1, 2013, for cohort one), which can be seen in the second column of Table II.C.2. (Table II.C.2 also serves a second purpose—to show the equivalence of the treatment and comparison practices at the start of the intervention—which we describe in Section II.C.4.). For benchmarking purposes, the last column shows the values of relevant variables for the national Medicare FFS population, when available.

Characteristics of the practices overall. Our analysis includes 18 cohort one treatment practices, seven of which are FQHCs or RHCs. Most treatment practices (72 percent) were located in an urban area and a primary care health shortage area. The 11 nonhealth center practices, on average, consisted of approximately four providers, with 95 percent of these providers having a primary care specialty. About a quarter of nonhealth center practices were owned by a hospital or health system (28 percent) or had providers who received payments from CMS for meaningful use of EHRs (22 percent) in the baseline period. None of the practices had any level of NCQA PCMH certification in the baseline period, consistent with the fact that a key aim of the WIPH PCMH intervention is to facilitate practices becoming NCQA-certified medical homes.

Characteristics of the practices' Medicare FFS beneficiaries. The characteristics of all Medicare FFS beneficiaries assigned to the treatment practices during the baseline period (January 1, 2012, through December 31, 2012) were similar to the nationwide FFS averages on some but not all characteristics. The average HCC risk score for the treatment group (0.95) was slightly lower than the national average (1.0). Patients in the treatment practices had hospital admission rates that were close to the national average. Medicare Parts A and B spending and the 30-day unplanned readmission rates were lower than the national average, but ED visit rates and inpatient admissions for ambulatory care-sensitive conditions were higher. The higher ED visit rate and admissions for ambulatory care-sensitive conditions might reflect the fact that these practices are serving a population in which primary care access might be limited, leading to higher ED and inpatient use.

4. Equivalence of the treatment and comparison groups at the start of the intervention

Demonstrating that the treatment and comparison groups were similar at the start of the intervention is critical for the evaluation design. This similarity increases the credibility of a key assumption underlying difference-in-differences models—that the change over time in outcomes for the comparison group is the same change that would have happened for the treatment group, had the treatment practices not received the intervention.

Table II.C.2 shows that the 18 treatment practices and the 69 selected comparison practices were similar at the start of the intervention on most matching variables. By construction, there were no differences between the two groups on whether the practice was a health center or a

Table II.C.2. Characteristics of treatment and comparison practices when the intervention began (January 1, 2013)

Characteristic of practice	Treatment practices (N = 18)	Unmatched comparison pool (N = 204)	Matched compar-ison group (N = 69)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Exact match variables[c]						
Characteristics of the practices overall						
Health center (%)	0.39	0.41	0.39	0	0	n.a.
Participating in the FQHC Demonstration Program (%)	0.06	0.02	0.06	0	0	n.a.
Certified as a PCMH by NCQA (%)	0	0[q]	0	0	0	n.a.
Propensity matched variables[d]						
Characteristics of the practice's location						
Located in an urban zip code (%)	72.2	56.3	70.3	2.0	0.04	n.a.
Located in a health professionals shortage area (primary care) (%)	72.2	44.3	68.0	4.2	0.09	n.a.
Characteristics of all patients attributed to practices during the baseline year (January 1, 2012 – December 31, 2012)						
Number of beneficiaries[e]	568.9	415.4	452.1	116.7	0.25	n.a.
HCC risk score	0.95	1.00	0.98	-0.03	-0.17	1.0[f]
All-cause inpatient admissions (#/1,000 patients/quarter)	72.68	65.75	72.33	0.35	0.01	74[g]
Outpatient ED visit rate (#/1,000 patients/quarter)	157.49	134.96	170.96	-13.46	-0.17	105[h]
Medicare Part A and B spending ($/patient/month)	768	674	735	33	0.14	860[i]
30-day unplanned hospital readmission rate (%)	12.4	10.6	12.0	0.4	0.07	16.0[j]
Inpatient admissions for ambulatory care-sensitive conditions (#/1,000 patients/quarter)	14.2	11.8	14.5	-0.33	-0.04	11.8[k]
Disability as original reason for Medicare entitlement (%)	21.5	23.8	22.4	-0.9	-0.08	16.7[i]
Dually eligible for Medicare and Medicaid (%)	17.5	16.2	18.3	-0.8	-0.08	22[p]
Age (years)	71.6	71.7	71.5	0.07	0.02	71[l]
Female (%)	55.0	55.8	55.3	-0.3	-0.04	55.3[g]
Characteristics of the practices (nonhealth centers only)[m]						
Providers in practice, according to SK&A (#)	3.9	4.8	4.1	-0.22	-0.08	n.a.
Providers in practice with primary care specialty (%)	94.5	91.5	94.9	-0.4	-0.03	n.a.
Owned by a hospital or health system (%)	27.8	27.1	25.6	2.2	0.05	n.a.
Meaningful use of EHRs (%)[n]	22.2	21.4	24.7	-2.5	-0.06	n.a.

Table II.C.2 *(continued)*

Characteristic of practice	Treatment practices (N = 18)	Unmatched comparison pool (N = 204)	Matched compar-ison group (N = 69)	Absolute difference[a]	Standard-ized difference[b]	Medicare FFS national average
Omnibus test for balance on matching variables[o]						
p-value		0.473 for health and nonhealth centers 0.352 for health centers only 0.436 for nonhealth centers only				n.a.

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS. Zip code information (whether an urban zip code or health professionals shortage area) was merged from the Area Resource File. Data on practices with NCQA recognition were merged from the NCQA database. Data on meaningful use of EHRs were merged from CMS.

Notes: The comparison group means are weighted based on the number of matched practices per treatment practice. For example, if four comparison practices are matched to one treatment practice, each of the four comparison practices has a matching weight of 0.25.

Absolute differences might not be exact due to rounding.

[a] The absolute difference is the difference in means between the matched treatment and comparison groups.
[b] The standardized difference is the difference in means between the matched treatment and comparison groups divided by the standard deviation of the variable, which is pooled across the matched treatment and matched comparison practices.
[c] Exact match means that a health center had to be matched to a health center and a nonhealth center had to be matched to a nonhealth center. We also exact matched health centers on whether they were participating in the FQHC demonstration program.
[d] We matched practices on these variables through propensity scores.
[e] We did not include the number of attributed beneficiaries in our propensity score model for nonhealth centers, but we did use this as a matching variable for health centers. We chose not to include this variable for matching nonhealth centers because we had differing data sources for the treatment and comparison practices on the number of providers working in these practices. Therefore for health centers, we matched on the number of providers working in practices, as counted through SK&A data, and not on the number of attributed beneficiaries. Because we explicitly did not match nonhealth centers on this variable, we accepted a standardized difference of 0.25, which was our maximum difference for balance.
[f] Chronic Conditions Data Warehouse (2014a).
[g] Health Indicators Warehouse (2014b).
[h] Gerhardt et al. (2014).
[i] Boards of Trustees (2013).
[j] Centers for Medicare & Medicaid Services (2014)
[k] This is the rate for all individuals ages 65 and older. Truven Health Analytics (2015).
[l] Health Indicators Warehouse (2014a).
[m] The 18 treatment practices include 11 nonhealth centers. There were 126 nonhealth centers in the unmatched comparison pool and 43 nonhealth centers in the matched comparison group.
[n] Meaningful use of EHRs is calculated as the percentage of practices with at least one provider (NPI) working in the practice who received financial incentives for meaningful use of certified EHRs through Medicare or Medicaid during the baseline period.
[o] Results from an overall chi-squared test indicate the likelihood of observing differences in the matching variables as large as the differences we observed if, in fact, the treatment and comparison populations (from which we drew the samples) were perfectly balanced. The value of $p = 0.473$ for the chi-squared test for the health centers and nonhealth centers suggests that the two groups are well balanced, because we cannot reject the null hypothesis that their characteristics are identical. For reasons described in the text, we excluded the number of attributed beneficiaries from the overall omnibus test on both types of practices but included it in the test for health centers only. The characteristics of the practices that were available for nonhealth centers only were included in the separate omnibus test for the nonhealth centers but not in the other omnibus tests. We also cannot reject the null hypothesis that the characteristics of the treatment and matched comparison groups are identical for the health centers only ($p = 0.352$) or for the nonhealth centers only ($p = 0.436$).
[p] Health Indicators Warehouse (2014c).
[q] As described in text, the potential comparison pool was limited to practices that did not have NCQA certification at the start of the intervention.
*/**/*** Significantly different from zero at the .10/.05/.01 levels, two-tailed test, respectively. No differences were significantly different from zero at the .10/.05/.01 levels.
CMS = Centers for Medicare & Medicaid Services; ED = emergency department; EHR = electronic health record; FFS = fee-for-service; FQHC = Federally Qualified Health Center; HCC = Hierarchical Condition Category; NCQA = National Committee for Quality Assurance; NPI = National Provider Identifier; PCMH = patient-centered medical home; PCP primary care provider; SD = standard deviation.
n.a. = not applicable.

nonhealth center; and whether the practice was participating in the CMS FQHC demonstration (applicable to health centers only). There were some differences between the treatment group practices and matched comparison practices on the variables included in the propensity-score model, but all the standardized differences across the propensity-score matching variables are within our target of 0.25 standardized differences, and most are within 0.15 standardized differences (the 0.25 target is an industry standard; for example, see Institute for Education Sciences 2014). The omnibus test that the treatment and comparison practices are perfectly matched on all variables common across both health and nonhealth centers cannot be rejected ($p = 0.47$), further supporting that the treatment and comparison groups were similar at the start of the intervention. Similarly, omnibus tests for the health center ($p = 0.35$) and nonhealth center ($p = 0.44$) subgroups were not statistically significant.

The differences for one variable, the number of attributed Medicare FFS beneficiaries, is exactly 0.25 standardized differences. On average, the treatment practices have more attributed Medicare FFS beneficiaries, overall (by 117 beneficiaries). However, as described earlier, we—in consultation with CMMI—decided not to require balance on this variable and we excluded it from the combined omnibus test for health and nonhealth centers.

5. Intervention impacts

In this section, we first present sample sizes and mean outcomes, by quarter, for the treatment and comparison groups. These mean outcomes provide context for understanding the difference-in-differences estimates; however, the differences in mean outcomes are not regression-adjusted and are not impact estimates by themselves. Next, we present results of the primary tests (which are regression-adjusted and averaged over the relevant period), by domain. Then we assess whether the primary test results are plausible given the secondary tests and the implementation evidence. We end with preliminary conclusions about program impacts in each domain.

a. Sample sizes

In the baseline period, the treatment group ranges from 8,896 (B1) to 11,537 (B4) beneficiaries (see Table II.C.3). The comparison group includes 29,589 to 32,931 unweighted beneficiaries during the same period. As expected, the sample for both the treatment and comparison groups grows steadily during each quarter of the baseline period. This happens because beneficiaries are newly assigned to practices over time, and beneficiaries who were previously assigned to the practices are retained in the sample unless they die, move out of the service areas, or become unobservable in FFS claims. The sample size for the treatment group drops between the last baseline quarter and the first intervention quarter from 11,537 to 10,968 beneficiaries (because beneficiaries no longer attributed to the treatment practices are dropped from the sample at that time). The sample then grows steadily again during the remaining seven intervention quarters to 14,794 beneficiaries for the same reason it grows in the baseline period. The comparison group follows the same pattern, with the unweighted sample ranging from 30,899 to 35,598 during the intervention period.

Table II.C.3. Sample sizes and unadjusted mean outcomes for Medicare FFS beneficiaries in the treatment and comparison groups for WIPH—PCMH Program, by quarter

	Number of Medicare FFS beneficiaries (practices)			Inpatient admissions for ambulatory care-sensitive conditions (#/1,000/quarter)			All-cause inpatient admissions (#/1,000/quarter)			Outpatient ED visit rate (#/1,000/quarter)			Medicare Part A and B spending ($/month)		
Q	T	C (not wgt)	C (wgt)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)	T	C	Diff (%)
Baseline period (January 1, 2012 – December 31, 2012)															
B1	8,896 (18)	29,589 (69)	9,398	16.7	15.9	0.9 (5.4%)	79.0	84.4	-5.4 (-6.4%)	153.5	185.7	-32.2 (-17.4%)	$769	$813	$-44 (-5.4%)
B2	9,823 (18)	30,878 (69)	9,901	15.1	15.9	-0.8 (-5.3%)	79.6	74.1	5.5 (7.4%)	154.5	217.6	-63.0 (-29.0%)	$812	$750	$62 (8.3%)
B3	10,702 (18)	31,931 (69)	10,489	13.0	19.4	-6.4 (-33.0%)	74.0	83.7	-9.7 (-11.6%)	149.6	207.5	-57.9 (-27.9%)	$773	$755	$18 (2.4%)
B4	11,537 (18)	32,931 (69)	11,171	14.9	14.1	0.8 (6.0%)	77.5	72.1	5.3 (7.4%)	154.7	184.8	-30.2 (-16.3%)	$810	$803	$7 (0.9%)
Intervention period (January 1, 2013 – December 31, 2014)															
I1	10,968 (18)	30,899 (69)	10,314	20.1	14.8	5.3 (35.8%)	84.4	80.2	4.3 (5.3%)	138.5	156.4	-18.0 (-11.5%)	$829	$754	$75 (10.0%)
I2	11,730 (18)	32,368 (69)	11,034	14.7	12.4	2.2 (18.0%)	72.4	69.2	3.1 (4.5%)	142.6	160.9	-18.3 (-11.4%)	$803	$707	$96 (13.6%)
I3	12,375 (18)	33,412 (69)	11,568	14.1	10.6	3.5 (32.9%)	75.5	65.2	10.3 (15.8%)	148.6	180.9	-32.3 (-17.9%)	$796	$720	$77 (10.7%)
I4	12,875 (18)	34,187 (69)	11,963	15.4	13.0	2.4 (18.7%)	75.3	61.5	13.7 (22.3%)	140.8	172.4	-31.6 (-18.3%)	$835	$714	$121 (16.9%)
I5	13,290 (18)	34,217 (69)	12,155	14.1	17.1	-2.9 (-17.2%)	73.4	70.8	2.7 (3.8%)	141.1	158.1	-17.0 (-10.8%)	$796	$688	$108 (15.7%)
I6	13,785 (18)	34,719 (69)	12,391	15.3	12.6	2.7 (21.4%)	79.5	69.6	9.9 (14.2%)	146.1	173.0	-26.9 (-15.6%)	$880	$789	$91 (11.6%)
I7	14,270 (18)	35,193 (69)	12,699	13.5	12.4	1.1 (8.7%)	72.5	69.1	3.4 (5.0%)	158.6	183.6	-24.9 (-13.6%)	$852	$754	$98 (13.0%)
I8	14,794 (18)	35,598 (69)	12,975	14.8	11.0	3.8 (35.0%)	70.2	64.6	5.7 (8.8%)	145.1	167.4	-22.2 (-13.3%)	$792	$716	$76 (10.6%)

Source: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Table II.C.3 *(continued)*

Note: The baseline quarters are measured relative to when the baseline period began on January 1, 2012. For example, the first baseline quarter (B1) runs from January 1, 2012, to March 31, 2012. The intervention quarters are measured relative to the start of the intervention period on January 1, 2013. The first intervention quarter (I1) runs from January 1, 2013, to March 31, 2013. The treatment and comparison group means are weighted such that (1) each treatment beneficiary gets a weight of one; and (2) each comparison beneficiary gets a weight that is the product of two weights: (a) a matching weight, equal to the reciprocal of the total number of comparison practices matched to the same treatment practice as the beneficiary's assigned practice, and (b) a practice size weight, which equals the average number of beneficiaries assigned to the matched treatment practice during the four baseline quarters divided by the average number of beneficiaries assigned to the beneficiary's comparison practice over those quarters. The difference between the treatment and comparison groups in a quarter is calculated by subtracting the mean outcome for the comparison group from the mean outcome for the treatment group. The percentage difference equals that difference divided by the mean outcome for the comparison group.

B = baseline; C = comparison; CMS = Centers for Medicare & Medicaid Services; Diff = difference; I = intervention; not wgt = not weighted; PCMH = patient centered medical home; Q = quarter; T = treatment; wgt = weighted; WIPH = Wyoming Institute of Population Health.

NA = not available.

n.a. = not applicable.

b. Mean outcomes for the treatment and comparison groups, by domain and quarter

Quality-of-care outcome. Inpatient admissions for ambulatory care-sensitive conditions were higher for the treatment group than the comparison group across most quarters (Table II.C.3).

Service use. The inpatient admission rates were higher for the treatment group than the comparison group during all intervention quarters. The ED visit rates for the treatment group were lower than the comparison group in all quarters.

Spending. Aside from B1, spending was higher across all quarters for the treatment group than the comparison group.

c. Results for primary tests, by domain

Overview. The primary tests are specified for the average impact of the intervention from I5 to I10 for cohort one and I5 and I6 for cohort two. For this report, we had data available only for I5 through I8 for cohort one. Thus, the primary tests in this report reflect the average impacts over only four intervention quarters (I5 through I8) for cohort one. Primary tests for all domains suggest substantively unfavorable effects (Table II.C.4). As described earlier, these results are preliminary because the analyses do not yet cover the full period or the cohort two practices that we will include in the final impact analysis in future reports.

Quality-of-care outcome. The treatment group's average number of inpatient admissions for ambulatory care-sensitive conditions was 14.4 per 1,000 beneficiaries per quarter during I5 through I8, which was estimated to be 1.1 more admissions than the counterfactual. (Our estimate of the counterfactual is the treatment group mean minus the difference-in-differences estimate.) This was an 8.3 percent unfavorable difference, which is greater than the substantive threshold of 5.0 percent. We cannot conclude that this is a statistically significant unfavorable effect because we used one-sided statistical tests. The statistical power values in Table II.C.4 imply that this analysis had limited power to detect differences in admissions for ambulatory care-sensitive conditions. The analyses only had a 16.3 percent power to detect a 5.0 percent difference in admissions for ambulatory care-sensitive conditions (the substantive threshold), and a 24.6 percent power to detect a 10.0 percent difference (twice the substantive threshold).

Service use. The treatment group's average number of all-cause inpatient admissions was 73.9 per 1,000 beneficiaries per quarter during I5 through I8. This was estimated to be 3.4 more admissions than the estimated counterfactual. This is a 4.8 percent unfavorable difference for all-cause admissions, which is greater than the substantive threshold of 3.75 percent. We adjusted for multiple statistical tests in this domain, but we cannot conclude that this is a statistically significant unfavorable effect because we used one-sided statistical tests. For ED visits, the treatment group averaged 147.7 visits per 1,000 beneficiaries per quarter during I5 through I8. This was estimated to be 17.3 more ED visits than the estimated counterfactual. This results in a 13.0 percent unfavorable difference, which is much higher than the substantive threshold of 5.0 percent. We also adjusted for multiple statistical tests in this domain for ED visits, but we cannot conclude that this is a statistically significant unfavorable effect because we used one-sided

Table II.C.4. Results of primary tests for WIPH—PCMH Program

Primary test definition				Statistical power to detect an effect that is[a]			Results			
Domain (# of test in domain)	Outcome (units)	Time period for impacts	Population	Substantive threshold (impact as a percentage of the adjusted comparison group mean)	Size of the substantive threshold	Twice the size of the substantive threshold	Treatment group mean	Regression-adjusted difference between treatment and estimated counterfactual (standard error)[b]	Percentage difference[c]	*p*-value[d]
Quality of care outcomes (1)	Inpatient admissions for ambulatory care sensitive conditions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 5–8	All observable Medicare FFS beneficiaries assigned to treatment practices[e]	-5.0%	16.3	24.6	14.4	1.1 (2.2)	8.3%	0.689
Service use (2)	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)			-3.75%	20.9	36.8	73.9	3.4 (5.6)	4.8%	0.611[f]
	Outpatient ED visits (#/1,000 beneficiaries/quarter)			-5.0%	14.7	20.6	147.7	17.3 (28.3)	13%	0.612[f]
	Combined (%)			-4.38%	16.4	25.1	n.a.	n.a.	9.0%	0.735[g]
Spending (1)	Medicare Part A and B spending ($/beneficiary/month)			-3.75%	24.0	44.7	$830	53.2 (50.7)	6.8%	0.853

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Note: The results for each outcome are based on a difference-in-differences regression model, as described in the text. Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

[a] The power calculation is based on actual standard errors from the analysis. For example, in the last row, a 3.75 percent effect on Medicare Part A and B spending (from the counterfactual of $830 -$53.20 = $776.80) would be a change of $29.13. Given the standard error of $50.70 from the regression model, we would be able to detect a statistically significant result 24.0 percent of the time if the impact was truly $29.13, assuming a one-sided statistical test at the $p = 0.10$ significance level.

[b] The counterfactual is the outcome the treatment group would have had in the absence of the HCIA-funded intervention. Our estimate of the counterfactual is the treatment group mean minus the regression-adjusted difference-in-differences estimate.

[c] Percentage difference is calculated as the regression-adjusted difference between the treatment and comparison groups, divided by the adjusted comparison group mean.

[d] *p*-values test the null hypothesis that the regression-adjusted difference-in-differences estimate is greater than or equal to zero (a one-sided test).

[e] Outcomes are observable for beneficiaries who are enrolled in Medicare FFS (Part A and B), are alive, and have Medicare as their primary payer.

f We adjusted the *p*-values from the primary test results for the multiple (two) comparisons made within the service use domain.

[g] This *p*-value tests the null hypothesis that the difference-in-differences estimates across the two outcomes in the service use domain, each expressed as percentage change from the comparison group mean, are greater than or equal to zero (a one-sided test).

*/**/*** Significantly different from zero at the .10/.05/.01 levels, one-tailed test, respectively.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; HCIA = Health Care Innovation Award; PCMH = patient-centered medical home; WIPH = Wyoming Institute of Population Health.

statistical tests. The mean percentage difference across all-cause inpatient admissions and ED visits was 9.0 percent (the average of 4.8 percent and 13.0 percent). This is a substantively unfavorable difference. Similar to the quality-of-care outcome, the service use outcomes had limited power to detect true impacts that were the size of the substantive threshold.

Spending. Medicare Part A and B spending per beneficiary per month averaged $830 for the treatment group during I5 through I8, which was estimated to be $53.2 higher than the estimated counterfactual. The 6.8 percent unfavorable difference is greater than the substantive threshold of 3.75 percent. As with the other outcomes, we cannot conclude that this is a statistically significant unfavorable effect. Power was also low for the spending outcome.

d. Results for secondary tests

Results from the secondary tests indicate unfavorable effects across all outcomes during I1 through I4 (Table II.C.5). Both the primary and secondary test results show that the outcomes for the comparison group were improving faster than those for the treatment group across all intervention quarters. A priori, we would have expected little to no change in outcomes for both treatment and comparison groups in the first year after the intervention started. The relatively early and large changes in outcomes among the comparison group in the intervention period suggests that we have to conduct additional analyses to understand the factors that might be driving our findings before we draw any conclusions about program effectiveness.

e. Consistency of quantitative estimates with implementation findings

As reported in Section II.A, WIPH's HCIA funds for the PCMH program were directed to TransforMED for practice facilitation services, to telehealth equipment, and to small grants to help pay NCQA application fees. Over time, the focus of the program shifted to NCQA application review. Although the PCMH program did not include an intensive intervention, it is unclear why patients in treatment practices would have such large unfavorable outcomes relative to patients in comparison practices as suggested by the quantitative estimates. Several practices participating in the PCMH program were working toward, and a few ultimately achieved, NCQA recognition during the HCIA award period, indicating that select practices were successfully implementing core elements of the PCMH model. On the other hand, a number of practices had difficulty with aspects of practice transformation, particularly related to EHR technology, which might have distracted them from optimal patient care or made it challenging to focus on other care improvements that would be expected to affect the outcomes examined.

Therefore, based on the implementation findings, it is yet to be determined whether the large unfavorable quantitative effects found in the first two years of the program are plausible given that (1) the intervention that was delivered was minimal and (2) the aspects of the program that practices *did adopt* during the intervention period would not be expected to produce substantive negative effects for the outcomes analyzed in the quantitative analysis. The two possible exceptions are if many practices were limited in their practice transformation and had reduced availability to treat patients due to difficulties with EHR adoption or if the practices that adopted core elements of the PCMH model were providing more comprehensive care that resulted in detection of additional health issues to address.

Table II.C.5. Results of secondary tests for WIPH—PCMH Program

Secondary test definition				Results			
Domain	Outcome (units)	Time period for impacts	Population	Treatment group mean	Regression-adjusted difference between treatment and estimated counterfactual (standard error)	Percentage difference	*p*-value
Quality of care outcomes	Inpatient admissions for ambulatory care sensitive conditions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 1 and 2	All observable Medicare FFS beneficiaries attributed to treatment practices	17.4	3.4 (2.6)	24.7%	0.910
		Average over intervention quarters 3 and 4		14.7	2.7 (2.6)	22.9%	0.854
Service use	All-cause inpatient admissions (#/1,000 beneficiaries/quarter)	Average over intervention quarters 1 and 2		78.4	1.6 (6.5)	2.1%	0.597
		Average over intervention quarters 3 and 4		75.4	9.9 (6.0)	15.1%	0.950
	Outpatient ED visits (#/1,000 beneficiaries/quarter)	Average over intervention quarters 1 and 2		140.5	21.1 (28.9)	17.6%	0.767
		Average over intervention quarters 3 and 4		144.7	7.0 (30.5)	5.1%	0.591
Spending	Medicare Part A and B spending ($/beneficiary/month)	Average over intervention quarters 1 and 2		$816	51.7 (54.9)	6.8%	0.827
		Average over intervention quarters 3 and 4		$816	61.0 (53.1)	8.1%	0.875

Sources: Analysis of the Medicare Enrollment Database and claims data accessed through the Virtual Research Data Center at CMS.

Note: The analyses in Table II.C.5 were conducted in the same way as the analyses in Table II.C.4.

*/**/*** Significantly different from zero at the .10/.05/.01 levels, one-tailed test, respectively.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; FFS = fee-for-service; PCMH = patient-centered medical home; WIPH = Wyoming Institute of Population Health

f. Conclusions about program impacts, by domain

Based on the preliminary evidence currently available, we are unable to draw any conclusions about program effectiveness (Table II.C.6). The primary tests in each of the three domains suggest that the program had substantively large unfavorable effects. However, more information is needed to determine whether these results are plausible given the secondary test results and the implementation evidence.

There are at least four possible reasons for why we might see large unfavorable difference-in-differences estimates. First, although the treatment and comparison practices were well matched on observable characteristics at baseline, there could have been unobserved differences between the groups or other confounding factors that influenced the results. Second, the estimates might be due to chance, particularly given the substantial statistical uncertainty in the estimates (as signaled by the low power to detect effects). Third, the composition of the sample might have been changing differentially in the treatment and comparison group due to attrition. Finally, it is possible that WIPH's other interventions influenced the composition of the treatment group in ways that made program impacts appear to be unfavorable. Specifically, one of the goals of WIPH's transitional care program is to connect patients to primary care. The transitional care program could have led, on average, to sicker beneficiaries (who had recently been hospitalized) to be assigned to the treatment practices (versus the comparison practices), making it appear that outcomes for the treatment group were worse than those for the comparison group.

g. Next steps for the impact evaluation

We plan to explore the data further and to conduct sensitivity tests to check the robustness of our estimates, focusing on the first, third, and fourth reasons described in Section II.C.5.f for why we might observe unfavorable impact estimates. Specifically, we will check how many of our matched comparison practices received NCQA PCMH recognition or payments for meaningful use of EHRs during the first year after the intervention started to assess whether practices in Montana were on a different trajectory of practice transformation and quality improvement that could not be detected at baseline and to determine whether the selected comparison group is a reasonable counterfactual. We will examine sample attrition over time to understand whether loss of sample to moving out of state or to death follows a different pattern in the treatment and comparison groups. We will test whether sicker beneficiaries were being assigned to the treatment practices (versus the comparison practices) in the intervention period because of the relationship of the program with WIPH's transitional care program. Lastly, we will also consider making follow-up calls to practices that were part of the intervention to better understand whether difficulties with EHR implementation might have affected patients' care for the outcomes examined.

As mentioned previously, the analyses do not yet cover the full period over which the program is expected to have effects or include the cohort two practices. In future reports, we plan to cover the full period and include the cohort two practices.

Table II.C.6. Preliminary conclusions about the impacts of WIPH—PCMH HCIA program on patients' outcomes, by domain

Domain	Preliminary conclusion	Evidence supporting conclusion		
		Primary test result(s)[a]	Primary test result plausible given secondary tests?	Primary test result plausible given implementation evidence?
Quality-of-care outcomes	None	Differences between treatment and comparison groups were substantively large and unfavorable for the single outcome in the domain	TBD	TBD
Service use	None	Differences between treatment and comparison groups were substantively large and unfavorable for both outcomes in the domain	TBD	TBD
Spending	None	Differences between treatment and comparison groups were substantively large and unfavorable for the single outcome in the domain	TBD	TBD

Sources: Tables II.C.4 and II.C.5

[a] More information is needed to determine whether the primary test results are plausible given the secondary test results and the implementation evidence. We will conduct additional analyses before making conclusions about program impacts.

PCMH = patient-centered medical home; TBD = to be determined; WIPH = Wyoming Institute of Population Health.

III. CONCLUSIONS AND NEXT STEPS FOR EVALUATION

WIPH received HCIA funding to create medical neighborhoods across Wyoming primarily through PCMH and WyRCT, the largest components of the award. The PCMH program provided training and facilitation to primary care practices to support PCMH transformation. WyRCT trained nurses to manage care transitions for high-risk patients discharged from acute care settings. WIPH also provided telehealth services, a medication donation program, and a virtual pharmacy program that used pharmacists to help eligible Medicaid patients with medication management. The multifaceted initiative aimed to reduce ED visits, hospital admissions, and total spending. The program was largely implemented on schedule, although the Virtual Pharmacy program experienced implementation delays and CMMI discontinued funding for that program in July 2014. Staff engagement facilitated the implementation process for PCMH and WyRCT. Some PCPs were engaged in TransforMED's learning collaboratives and worked with TransforMED consultants to help prepare PCMH applications for NCQA recognition. Conversely, lack of provider engagement was a program barrier for some PCMH practices, especially those that did not have a designated physician champion to advocate for and lead the transformation. The implementation of WyRCT also benefited from having highly engaged nurses conduct home visits and provide coaching and mentoring to program participants. The HCIA-Primary Care Redesign Clinician Survey found that most clinicians believed the HCIA-funded initiative would have a positive effect on patient-centeredness. About one-third of clinicians surveyed believed it would have a positive effect on patients' safety, quality of care, and clinicians' ability to respond in a timely way to patients' needs.

Based on the preliminary evidence currently available for the PCMH program, we are unable to draw any conclusions about program impacts on patient outcomes in the three domains examined: quality of care, service use, or medical spending. The primary tests found large unfavorable differences between the treatment and comparison groups in all domains for Medicare FFS beneficiaries during the primary test period (months 13 through 24 after the program began). However, more information is needed to determine whether these primary results are plausible. The implementation evidence collected thus far does not provide any clear rationale for why the PCMH program would have consistently unfavorable impacts. Further, the secondary tests—which found large unfavorable differences in the first year of practice participation, when no or only small effects were expected—suggest there could be limitations in the matched comparison group, such as differential addition or attrition to the comparison versus treatment group which could cause differences in mean outcomes unrelated to program impacts. Further, we have yet to include the full primary test period (months 13 through 30 after the program began) and cohort two practices.

Our next steps for this evaluation are to (1) monitor WIPH's program implementation reports through June 30, 2015, and plans for sustaining the program beyond the funding period; (2) evaluate trainees' and clinicians' attitudes toward and experiences with the program in the third year of the award through administered surveys; (3) conduct the robustness checks for the WIPH-PCMH impact evaluation described in Section II.C.5.g; (4) extend the WIPH-PCMH impact evaluation to include the full period of program operations and the second cohort of

practices; (5) include results from the impact evaluation of WIPH's care transitions program; and (6) use the implementation findings to help interpret the impact results.

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

REFERENCES

American Diabetes Association. "Standards of Medical Care in Diabetes – 2015." *The Journal of Clinical and Applied Research and Education,* vol. 38, no. 1, 2015, p. 534. Available at http://professional.diabetes.org/admin/UserFiles/0%20-%20Sean/Documents/January%20Supplement%20Combined_Final.pdf. Accessed June 24, 2015.

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. "2013 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." Table V.D1. Washington, DC: Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, 2013. Available at http://downloads.cms.gov/files/TR2013.pdf. Accessed August 13, 2014.

Centers for Medicare & Medicaid Services. "Chronic Conditions Among Medicare Beneficiaries, Chartbook, 2012 Edition." Baltimore, MD: U.S. Department of Health and Human Services, 2012. Available at http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/2012Chartbook.pdf. Accessed February 23, 2015.

Centers for Medicare & Medicaid Services. "CSV Flat Files—Revised: Readmissions Complications and Deaths—National.csv." Baltimore, MD: CMS, 2014. Available at https://data.medicare.gov/data/hospital-compare. Accessed August 14, 2014.

Chronic Conditions Data Warehouse. "Table A.1. Medicare Beneficiary Counts for 2003 – 2012." Baltimore, MD: Centers for Medicare & Medicaid Services, 2014a. Available at https://www.ccwdata.org/cs/groups/public/documents/document/ccw_website_table_a1.pdf. Accessed November 19, 2014.

DesRoches, C.M., M.W. Painter, and A.K. Jha. "Health Information Technology in the United States: Progress and Challenges Ahead, 2014." Princeton, NJ: The Robert Wood Johnson Foundation, 2014.

Furukawa, M.F., J. King, V. Patel, C. Hsaio, J. Adler-Milstein, and A.K. Jha. "Despite Substantial Progress in EHR Adoption, Health Information Exchange and Patient Engagement Remain Low." *Health Affairs,* vol. 33, no. 9, 2014, pp. 1672–1679.

Gerhardt, Geoffrey, Alshadye Yemane, Keri Apostle, Allison Oelschlaeger, Eric Rollins, and Niall Brennan. "Evaluating Whether Changes in Utilization of Hospital Outpatient Services Contributed to Lower Medicare Readmission Rate." *Medicare & Medicaid Research Review,* vol. 4, no. 1, 2014, pp. E1–E13.

Health Indicators Warehouse. "Average Age of Medicare Beneficiaries." Hyattsville, MD: National Center for Health Statistics, HIW, 2014a. Available at http://www.healthindicators.gov/Indicators/Average-age-of-Medicare-beneficiaries-mean_308/Profile/ClassicData. Accessed November 19, 2014.

Health Indicators Warehouse. "Hospital Inpatient Medicare Admissions (per 1,000 beneficiaries)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014b. Available at http://www.healthindicators.gov/Indicators/Hospital-inpatient-Medicare-admissions-per-1000-beneficiaries_2001/Profile/ClassicData. Accessed August 13, 2014.

Health Indicators Warehouse. "Medicare Beneficiaries Eligible for Medicaid (percent)." Hyattsville, MD: National Center for Health Statistics, HIW, 2014c. Available at http://www.healthindicators.gov/Indicators/Medicare-beneficiaries-eligible-for-Medicaid-percent_317/Profile/ClassicData. Accessed August 4, 2015.

Institute of Education Sciences. "What Works Clearinghouse: Procedures and Standards Handbook, Version 3.0." Washington, DC: U.S. Department of Education, IES, 2014. Available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19. Accessed September 15, 2014.

Peikes, Deborah, Stacy Dale, Eric Lundquist, Janice Genevro, and David Myers. "Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper." AHRQ Publication No.11-0100-EF. Rockville, MD: Agency for Healthcare Research and Quality, October 2011.

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science,* vol. 25, no. 1, 2010, pp. 1–21.

Truven Health Analytics. *AHRQ Quality Indicators, Prevention Quality Indicators v5.0 Benchmark Data Tables*. Prepared for the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. Santa Barbara, CA: Truven Health Analytics, March 2015. Available at http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V50/Version_50_Benchmark_Tables_PQI.pdf. Accessed August 18, 2015.

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research