MATHEMATICA
Policy Research

MEDICAID
1115
DEMONSTRATIONS

Medicaid 1115
Demonstration
Evaluation
Design Plan

Delivery System Reform Incentive Payments

Design Supplement:
Interim Outcomes Evaluation
June 2017

Julia Baller

Marian V. Wrobel

Natalya Verbitsky Savitz

Mariel Finucane

Carol Irvin

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

CONTENTS

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

This document is a supplement to the Medicaid 1115 Demonstration Evaluation Design Plan prepared by Mathematica Policy Research and submitted to the Centers for Medicare & Medicaid Services (CMS) in May 2015 (Irvin et al 2015). The supplement presents our approach to the interim evaluation of the section 1115 Delivery System Reform Incentive Payment (DSRIP) demonstrations. The purpose of the interim evaluation, which we will conduct in 2017, is to examine the preliminary effects of the DSRIP demonstrations on delivery system transformation and on specific clinical quality outcomes. The results of this evaluation will provide useful information about the short-term effects of the demonstrations on key outcomes and on the settings in which the DSRIP has been most effective. The final evaluation in 2019 will assess the more mature impacts of the demonstration on provider readiness for value-based purchasing, on population health, and on the total cost of care.

The sections that follow describe the DSRIP demonstrations and the methods we intend to use to evaluate them, the research questions, outcome measures, eligible populations, comparison groups, data sources, and analytic strategies.

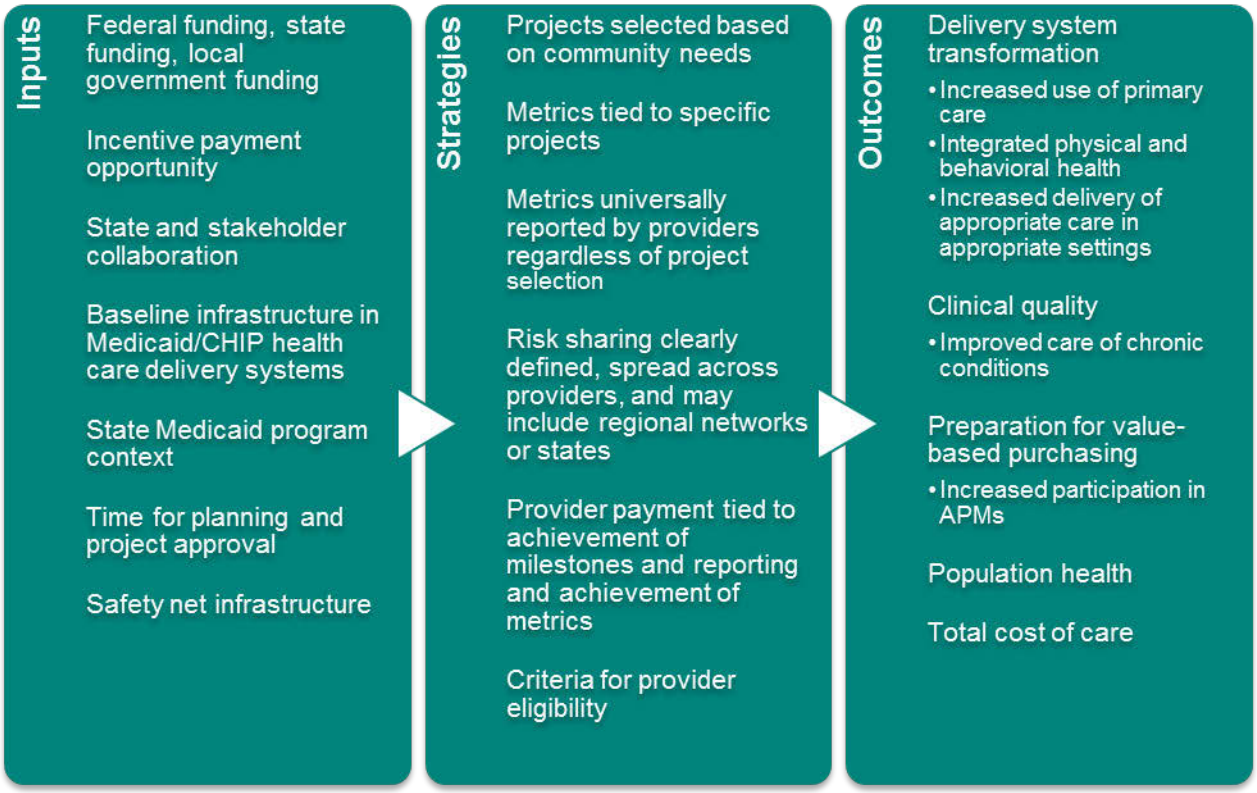
A. DSRIP demonstrations

The interim evaluation will focus on demonstration programs in California, New Jersey, and Texas, the three states in which the program has been operating for long enough and that have enough data to support a meaningful study.[1] The DSRIP demonstrations in these states provide federal funding to providers so that they can conduct projects that seek to transform the delivery system and thereby improve the quality of care, improve patient outcomes, reduce the cost of care, and prepare for value-based purchasing (**Figure 1)**.

Although the DSRIP demonstrations have the same broad goals and operational framework, they vary considerably in other respects across the study states (see Appendix Table A.1. for DSRIP demonstration characteristics by state). For example, early DSRIP demonstrations in California and Massachusetts were primarily intended to provide financial support for safety net health systems that serve a high volume of Medicaid beneficiaries and the uninsured. In larger demonstrations in Texas and New York, there is more emphasis on transforming the delivery system across care settings and provider organizations to improve population health. In addition to variation across states, there is considerable variation across providers within a state with regard to the number and types of projects that are being implemented and the number and types of milestones and measures being reported (see Appendix A.2 for details about the project selected by providers). These sources of variation play a critical role in how we designed the interim evaluation.

[1] As of May 2017, there are DSRIP programs in California, Massachusetts, Texas, New Hampshire, New Jersey, New York, Rhode Island, and Washington. The final evaluation will include as many of these states as possible. However, the interim evaluation relies on claims data, which are only available through certain years for certain states. The interim evaluation is therefore limited to states in which claims data are available for at least one year after the DSRIP program was implemented.

Figure 1. Logic model for the DSRIP demonstration



B. Research questions and overview of the interim evaluation

The intent of the DSRIP program is to broadly affect how care is delivered to Medicaid beneficiaries and the uninsured. There may also be other related effects, such as bolstering the finances of safety net hospitals. The interim evaluation focuses on a limited set of measures related to the quality and efficiency of care that are intended to serve as sentinel indicators not only for system transformation because care is delivered differently, but also for population health because high quality care should improve population health. In the final evaluation, if data are available, we will examine a broader set of quality and efficiency measures, and will add measures that assess providers' readiness for value-based purchasing, population health outcomes, and the total cost of care.

The interim evaluation will address two over-arching research questions:

1. What was the overall effect of DSRIP demonstrations on key outcomes related to delivery system transformation and clinical quality?
2. Under what circumstances are DSRIP demonstrations more or less effective?[2]

The analysis will focus on four clinical outcome measures that reflect the demonstration's overall purpose of transforming care and that are likely to respond relatively quickly to

[2] Effectiveness is defined by statistically significant relative changes in the level or trends in the outcomes of interest.

demonstration projects: (1) emergency department (ED) visits; (2) follow-up after ED discharge for patients with selected chronic medical conditions (asthma, chronic obstructive pulmonary disease [COPD], hypertension, or diabetes); (3) follow-up after ED discharge for patients with mental illness; and (4) hemoglobin A1c testing (**Table 1)**.

Table 1. Outcome measures

Outcome measure	NQF number	In a Medicaid core set	Clinical focus area(s)	Numerator	Denominator
ED visits	NA	Yes[a]	Primary care Appropriate care in appropriate settings	ED visits not resulting in an inpatient admission	Enrollee months for adults ages 18 to 64
Follow-up after discharge from the ED for asthma, COPD, hypertension, and diabetes	NA	No	Primary care	Eligible adults with an outpatient visit within 7 days of discharge from the ED for asthma, COPD, hypertension and diabetes	Adults ages 18 to 64 with an ED visit for asthma, COPD, hypertension, or diabetes[b]
Follow-up after discharge from the ED for mental illness	NQF 2605	No	Physical and mental health integration	Eligible adults with an outpatient visit within 7 days of ED discharge for mental illness	Adults ages 18 to 64 with an ED visit for mental illness[b]
Comprehensive diabetes care: Hemoglobin A1c control	NQF 0059	Yes	Diabetes care	Eligible adults with HbA1c testing	Adults ages 18 to 64 with diabetes [c]

[a]The Medicaid core set of child quality measures includes the measure Ambulatory Care—ED visits for beneficiaries ages 0 to 21. We have adapted this measure for the adult beneficiaries ages 18 to 64. Most notably, the measure excludes ED visits for mental illness, and alcohol and other drug dependence, but we will include these visits in this measure for adults when possible.

[b]Denominator excludes ED visits that result in an inpatient admission.

[c]Denominator includes beneficiaries who have been diagnosed with diabetes in the measurement year or any year prior.

COPD = chronic obstructive pulmonary disease; ED = emergency department; NA = not applicable? not available? NQF = National Quality Forum

The preferred analytic approach for the interim evaluation will be a comparative interrupted time series. We will use data from the Medicaid Analytic eXtract (MAX) to compare, before and after the demonstration was implemented, patient-level outcomes and their trends for individuals living in hospital service areas (HSAs) served by DSRIP providers with outcomes and trends for individuals living in similar HSAs that are not served by DSRIP providers,. When a suitable comparison group is not available, we will use a simple interrupted time series in which we examine changes in both the level and the trend of patient-level outcomes before and after the demonstration was implemented.

C. Outcome measures

As mentioned, the analysis will focus on four outcome measures that are intended to capture a fundamental shift toward primary care and improved care coordination, leading to declines in avoidable hospital use. In selecting these measures, we sought to reflect CMS's and state priorities for their DSRIP demonstrations, to include measures relevant to the most common clinical focus areas of the projects,[3] and to use endorsed measures and measures used by key project stakeholders when possible. We also sought to analyze a small number of measures to make the effort more focused, to ensure that the findings are clear and easy to understand, to avoid the loss of statistical power because of multiple comparisons, and to design an analysis that was feasible given the time and resources available.

ED visits. We will measure the rate of ED visits per 1,000 enrollee months among adults ages 18 to 64. To our knowledge, no measures of availability or the use of primary care among adults are currently endorsed.[4] As such, we propose to apply the ED visits measure in the Medicaid core set of child quality measures to the adult population as a proxy that represents lack of access to primary care (CMS 2016). If the DSRIP demonstrations increase access to primary care services, the use of the ED should decline.

Follow-up after discharge from the ED for ambulatory care sensitive conditions (asthma, COPD, hypertension, and diabetes). We will examine follow-up after discharge from the ED for ambulatory care sensitive conditions as another measure of access to primary care. We will measure the rate of follow-up within seven days of discharge from the ED for asthma, COPD, hypertension, and diabetes for visits that do not result in an inpatient admission. Standards for high quality care indicate that many patients who visit the ED for these conditions should have a primary care visit soon afterwards.[5] More generally, individuals who do not receive follow-up care are more likely to be readmitted to the ED (Cook et al. 2004).

Follow-up after discharge from the ED for mental illness. As we propose for ambulatory care sensitive conditions, we will examine follow-up after discharge from the ED for mental health conditions within seven days after discharge to assess the extent to which the DSRIP is more fully integrating physical and mental health care and raising the quality of mental health services. Individuals with mental health conditions are particularly vulnerable to losing contact with the health care system after an ED visit. Furthermore, DSRIP providers commonly implement projects focused on integrating care for individuals with mental illness.

[3] To better understand state and provider clinical priorities, we developed a taxonomy of clinical focus areas that was streamlined, comprehensive, and reflective of the key goals of the DSRIP demonstrations. We mapped each project to one or more of the clinical focus areas. See Appendix Tables A.1 and A.2 for common areas of clinical focus and the extent to which each state adopted projects in these areas.

[4] To identify measures of availability and use of primary care among adults, we reviewed administrative measures endorsed by the National Quality Forum (NQF). The team was not able to identify any NQF-endorsed administrative measures of access to, or the use of, primary care by adults. We then reviewed the measures database we developed under this contract to identify primary care measures for adults reported by DSRIP providers. We found that there are no consistent measures of adult primary care reported across the DSRIP demonstrations.

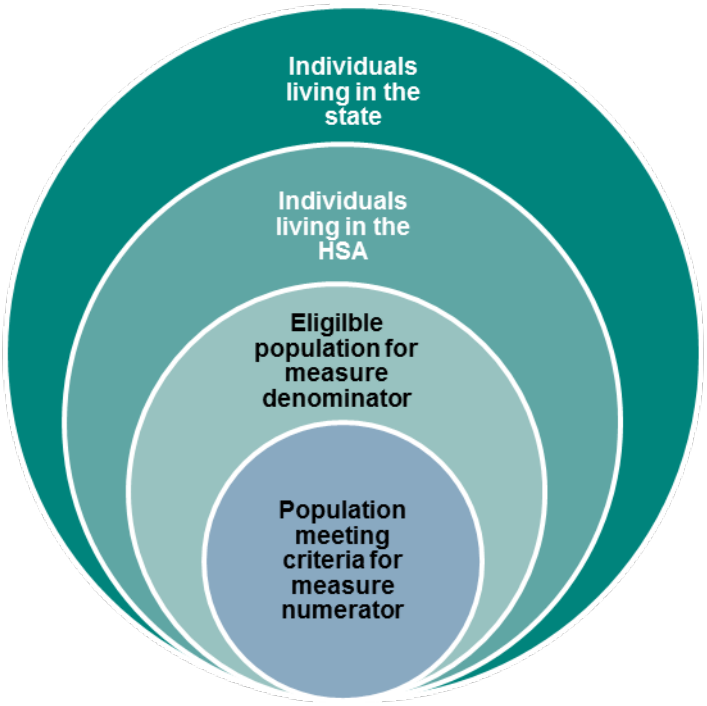
[5] Based on discussions with Mathematica's clinical experts.

Comprehensive diabetes care: HbA1c testing. Finally, we will measure HbA1c testing to assess whether DSRIP demonstrations are influencing the quality of diabetes care. Diabetes is a condition that is highly prevalent among Medicaid beneficiaries, and DSRIP providers commonly select projects that focus on improving care for beneficiaries with diabetes. This measure is endorsed by the NQF and part of the Medicaid Core Set of Adult Health Care Quality Measures (CMS 2016).

D. Eligible population for measure calculation

The DSRIP demonstrations are intended to affect care for the entire community and across a spectrum of providers. To reflect this, we will define the population eligible for the demonstration as all continuously enrolled, full-benefit Medicaid beneficiaries who reside within the catchment area of participating hospitals. We will use the Dartmouth Atlas hospital service areas (HSAs) to define the hospital catchment areas (Dartmouth Institute for Health Policy and Clinical Practice 2017). We will then define the denominators and numerators for each measure within that population, as illustrated in **Figure 2.** In keeping with the focus of the DSRIP program and the definitions of our outcome measures, the sample will consist of adults ages 18 to 64 who are not disabled.

Figure 2. Eligible population for measure calculation



E. Selecting appropriate comparisons

For the purpose of this evaluation, the effect of the demonstration will be the difference between the observed outcomes in participating communities and the outcomes that would have occurred in those communities if the DSRIP program had not been implemented (the counterfactual). To estimate this effect, we will use a comparative interrupted time series design to examine whether the outcomes in the demonstration group deviate from baseline trends to a

greater extent than do outcomes in the comparison group during the demonstration period. A simple interrupted time series design looks for a deviation from baseline trends during the demonstration period without using a comparison group.

Given the differences in the states' DSRIP programs, we will select the analytic design and construct comparison groups separately for each state by using the same framework (see **Figure 3**). Given the important differences between states in policy factors such as what services or populations are covered by Medicaid and the extent of cost-sharing, the preferred analytic design will be a comparative interrupted time series with an in-state comparison group. If this approach is not feasible, we will use a comparative interrupted time series with an out-of-state comparison group, followed by a simple interrupted time series.

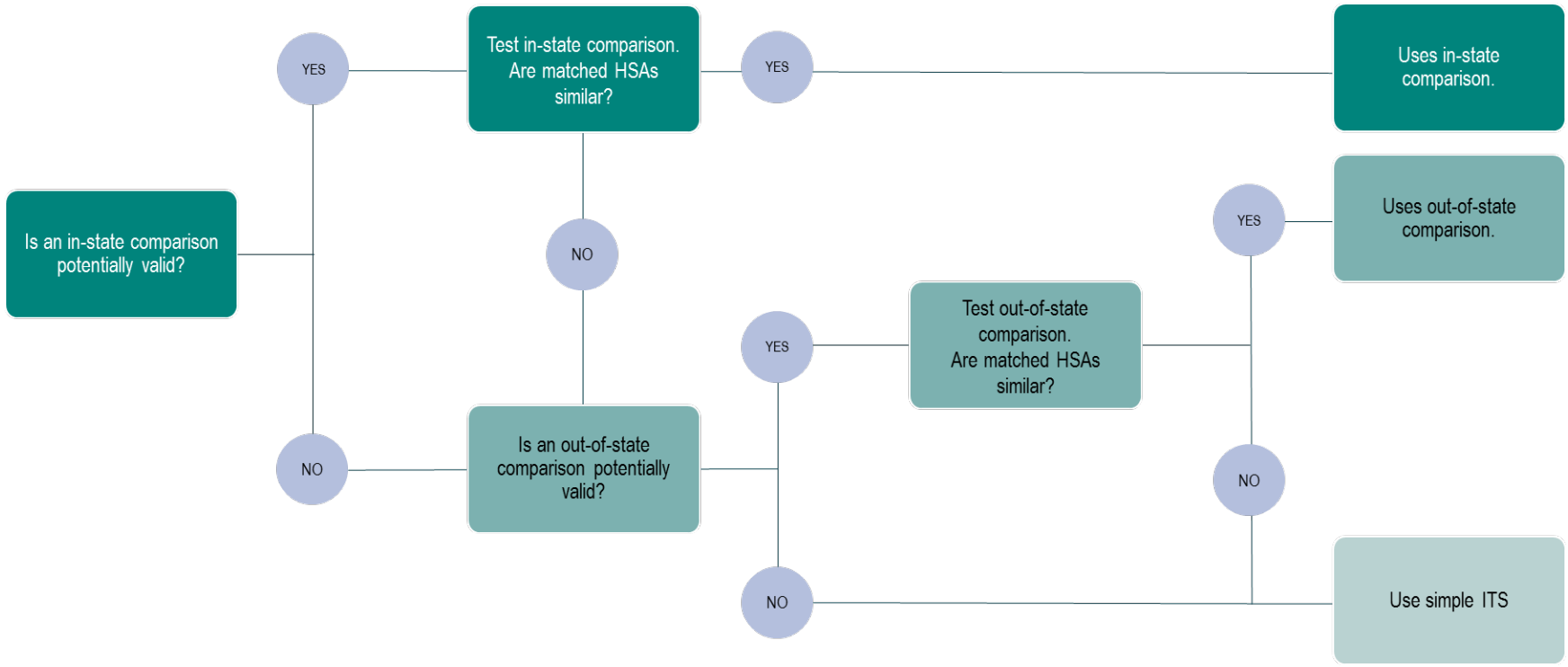
More specifically, for each state, we will determine whether the process by which demonstration providers were selected suggests that an in-state comparison group is potentially valid. For example, all New Jersey acute care hospitals were eligible for the DSRIP, and some explicitly opted out, suggesting that participating and nonparticipating hospitals are likely to differ in their commitment to reform. We will then determine whether there are enough HSAs within the state to create a comparison group. If a within-state comparison group is not a viable option, we will look for an out-of-state comparison group. We will begin with a set of states identified by CMS as places where state policymakers and providers are engaging in or preparing for similar delivery system reform.[6] We will then determine whether one or more of these states is similar to the demonstration state in terms of basic regional, economic, and demographic characteristics. If no in-state or out-of-state comparison group is available, we will default to a simple interrupted time series design for the interim evaluation.[7]

If we can identify an appropriate comparison group (either within-state or out-of-state), we will select comparison HSAs by matching to demonstration HSAs on four variables. Three of these variables reflect the distinctive characteristics of DSRIP hospitals—percentage of discharges with Medicaid as payer, hospital size, rural or urban location—whereas one is related to the outcomes of interest: the HSA's rate of ambulatory care sensitive Medicare discharges in the baseline period. We will then test whether the resulting comparison group is similar to the demonstration group in terms of other characteristics such as percent uninsured in the HSA, community median income in the HSA, the rates of chronic conditions among patients, and the levels and trends of outcome measures in the pre-period.

[6] These states are Maryland, Oregon, Virginia, and Washington. New York is also considered, as it had not yet implemented its DSRIP demonstration during the study period.

[7] For the final evaluation, we will reassess the potential for creating out-of-state comparison groups, drawing, perhaps, on a wider set of candidate states.

Figure 3. Comparison group selection



It is likely that this process will lead to different results for each demonstration state. In California, where all 21 designated public hospital systems were eligible for the DSRIP, we will explore creating a within-state comparison group. In New Jersey, where almost all hospitals participate in the DSRIP, we will explore an out-of-state comparison group of New York HSAs. Finally, in Texas, we will rely on a simple interrupted time series design because the DSRIP is being implemented across the state, and Texas differs too greatly from all of the candidate comparison states in terms of region, percentage of residents uninsured, percentage of residents with low incomes, percentage of residents living in rural areas, and the age distribution of the population.

In all states, findings regarding the effect of the DSRIP program will be complemented by findings regarding the effect of the level of DSRIP funding, an analysis that relies on cross-sectional variation between HSAs participating in the DSRIP. Estimates that rely on such cross-sectional variation are not biased by differences between DSRIP and non-DSRIP hospitals or unmeasured variables that change over time, so they serve as a robustness check for the interrupted time series estimates.

F. Data sources

Medicaid enrollment and claims data. The primary data source for assessing the outcome measures will be the Medicaid Statistical Information System (MSIS) research files, known as the Medicaid Analytic eXtract (MAX) files and their early versions (Alpha-MAX).[8] The MAX data offer a comprehensive enrollment and claims history for each Medicaid beneficiary, enabling us to study outcomes at the individual level and control for demographic and clinical covariates. These data are available from 2009 through 2014 for California, New Jersey, and New York and through 2013 in Massachusetts and Texas (Table 2). Given the limited availability and quality of the MAX data in Massachusetts, we excluded the state from the interim outcomes evaluation.[9]

Data limitations in each state, described in Table 2, influenced our selection of outcome measures, and they will likely impose some additional limitations on measure construction. For instance, because California and Texas do not have usable inpatient encounter records for adult beneficiaries at some points during our study period, we selected only measures that rely on outpatient data. In addition, the known limitations of behavioral health organization (BHO) encounter data (Nysenbaum et al. 2014) may limit our ability to calculate measures that are based on outpatient behavioral health visits in California.

[8] CMS develops MAX data as a more research-friendly version of MSIS files. MAX production requires seven quarters of MSIS data, including four quarters for the calendar year plus three additional quarters with adjustment records. Alpha-MAX data are produced without the full seven quarters of MSIS data.

[9] Mathematica does plan to include Massachusetts in the final evaluation, drawing upon data from the state's hospital discharge database.

Table 2. MAX data availability

State	DSRIP demonstration approval date	Implementation start date	Data availability	Quarters of data post-implementation	Include in interim impact evaluation
California	2010	2011	Through 2014[a]	12	Yes
Massachusetts	2011	2012	Through 2013[b]	6	No
New Jersey	2012	2014	Through 2014	4	Yes
Texas	2011	2011	Through 2013[c]	9	Yes
New York	2014	2015	Through 2014[d]	0	Yes, as a comparison state

[a]California has no usable inpatient encounter data for child, disabled, and aged populations from 2009 to 2011, and there are none for adults in 2011. In addition, the state provides behavioral health services through behavioral health organizations, many of which report incomplete data.

[b]Massachusetts has no inpatient or ambulatory care encounter records from 2009 to 2010, no inpatient encounter records in 2011, and no ambulatory care encounter records for the aged population in 2011.

[c]Texas has no usable inpatient encounter records from 2009 to 2011 for the adult, disabled, and aged population; no usable ambulatory care encounter records for the aged population from 2012 to 2013; and no inpatient encounter record procedure codes in 2013.

[d]New York has no usable inpatient encounter data for children from 2010 to 2011 and no usable ambulatory care encounter data for children in 2011.

Additional data sources. In addition to the MAX data, we are using the following data sources to construct variables that we will use in creating matched comparison groups, estimating overall program effects, and conducting subgroup analyses.

- DSRIP state and provider documentation to construct variables that capture demonstration characteristics (e.g., eligible incentive payments by provider or project selection by providers)
- American Hospital Association annual survey of hospitals to define hospital-level characteristics (e.g., the number of beds)[10]
- American Community Survey to define zip-code–level sociodemographic characteristics (e.g., median household income)[11]

[10] All measures constructed from the American Hospital Association annual survey of hospitals will be based on the 2009 survey.

[11] All measures constructed from the American Community Survey will be based on the 2013 five-year estimates.

- Publicly available data on health professional shortage areas to define areas in which there is a shortage of primary and mental health care[12]

- Medicare Healthcare Cost Reporting System data to determine the share of hospital discharges that are paid for by Medicaid in each HSA[13]

G. Estimating demonstration effects

To estimate the comparative or simple interrupted time series models with the MAX data, we will use regression models in which the unit of observation is the person-quarter, and the dependent variables are the clinical outcomes.[14] As noted, we will estimate these models separately by state, given the differences between states in program design, in the availability of a comparison group, and in the data.

1. What was the overall effect of the DSRIP on key outcomes?

To address the first research question in California and New Jersey—assuming we can identify an appropriate comparison group—we will estimate two patient-level regression models, one for each state, modeling an outcome (*y)* for person (*i*), in HSA (*j*) at time (*t):*

$$y_{ijt} = \beta_0 + \beta_1 \mathrm{T}_{ij} + \beta_2 \mathrm{X}_{ijt} + \beta_3 \mathrm{T}_{ij} \mathrm{X}_{ijt} + \beta_4 \mathrm{Z}_{ij} + \beta_5 \mathrm{Z}_{ij} \mathrm{T}_{ij} + \beta_6 \mathrm{Z}_{ij} \mathrm{X}_{ijt} + \beta_7 \mathrm{Z}_{ij} \mathrm{X}_{ijt} + \beta_8 W_{ij} + \alpha_i + b_j + \varepsilon_{ijt}$$

This model includes four types of covariates:

- T_{ij} is a time covariate
- X_{ijt} is an intervention indicator, equal to 1 if the observation is in the post-period and equal to 0 if the observation is in the pre-period
- Z_{ij} is a treatment indicator, equal to 1 if the HSA *j* is affected by the DSRIP and equal to 0 if HSA *j* is in the comparison group
- W_{ij} contains patient-level characteristics such as age, gender, presence of chronic conditions; characteristics of the patient's home zip code such as median income and whether it is classified as rural, a primary care shortage area, or a behavioral health shortage area; and characteristics of the HSA such as percentage of residents uninsured or covered by Medicaid, hospital beds per resident, and number of hospitals

The model includes three error terms:

- α_{ij} is a patient-level random effect
- b_j is a HSA-level random effect

[12] All measures constructed from the health shortage areas will be based on 2013 data.

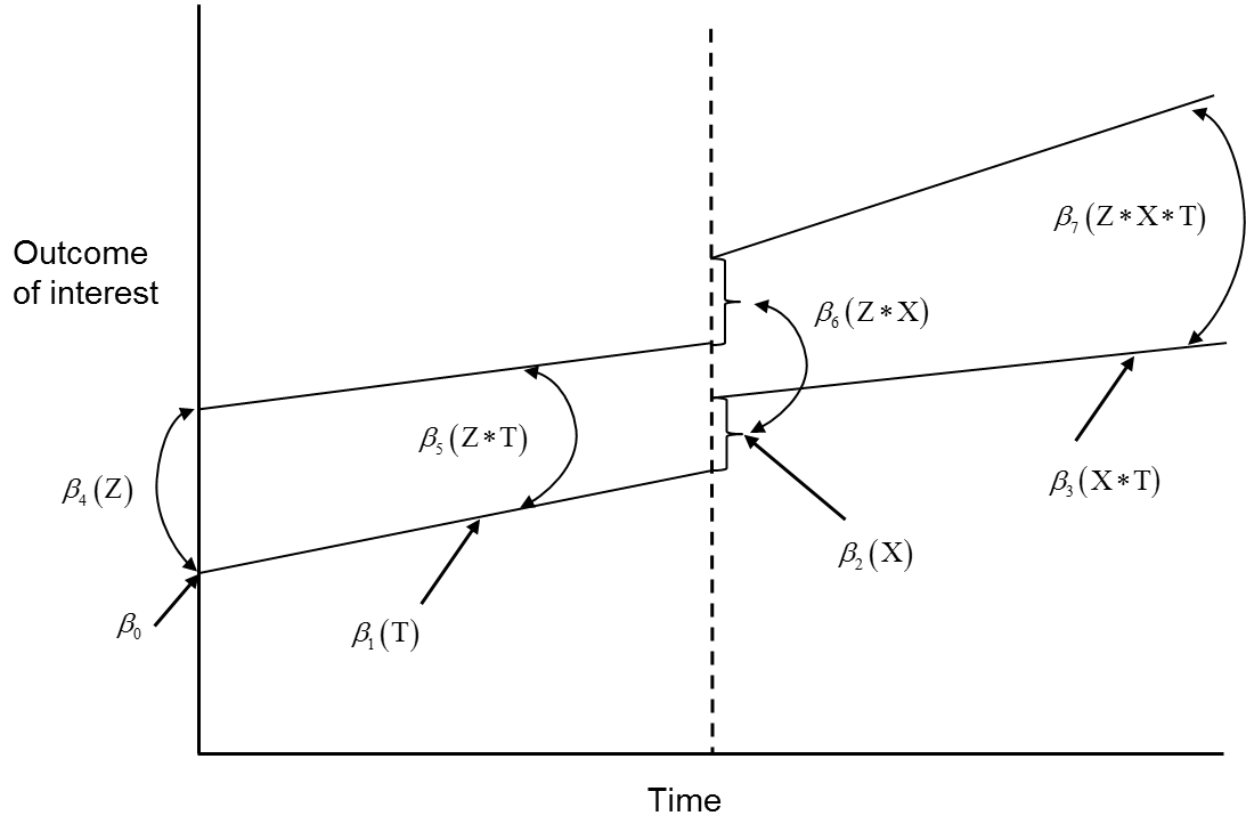
[13] All measures constructed from the Medicare Healthcare Cost Reporting System will be based on 2009 data.

[14] For the diabetes measure, the unit of observation is the person-year, which corresponds to the technical specifications in the Medicaid adult core set of clinical quality measures.

- ε_{ijt} is a residual error term[15]

The coefficients of the model are shown in **Figure 4** below. The estimate β_6 reflects the differential impact of the demonstration, relative to the comparison, on the *level* of the outcome variable, and the estimate β_7 reflects the differential impact of the demonstration on the *trend* of the outcome variable. Taken together, these two coefficients represent the estimate of the demonstration effect.

Figure 4. Illustration of the comparative ITS



The hierarchical error structure of this model accounts for unmeasured differences between patients and HSAs, and the nested nature of the data (i.e., patients are nested within HSAs). In other words, the model also assumes that individuals within each HSA are more similar to each other than to individuals in other HSAs.

In Texas, we will use a simple interrupted time series design to estimate whether the level or trends in the outcomes of interest in the pre-demonstration period are significantly different from the outcomes of interest in the demonstration period. In this simplified model, we will include covariates for time; an intervention indicator equal to one if the observation is in the post-period

[15] We will also explore including random slopes in the comparative interrupted time series models to allow for the possibility that trends in the outcomes of interest vary by HSA.

and equal to zero if the observation is in the pre-period; and patient–, zip-code–, and HSA-level covariates measured at baseline.

2. Under what circumstances are the DSRIP demonstrations affecting key outcomes?

In addition to quantifying the overall effect of the DSRIP, it is also important to understand the effects of particular DSRIP features and the contexts in which the demonstration operates so that this type of demonstration can be refined and replicated. For example, we may be interested in understanding whether hospitals that are implementing diabetes projects improve the quality of diabetes care more rapidly than hospitals that are not doing so, or whether the impact of the DSRIP was greater for patients with behavioral health conditions. In addition to informing CMS about the relative importance of different features and contexts, this analysis will also ensure that we understand whether certain characteristics of DSRIP demonstrations have an effect on outcomes, even if DSRIP is found to have no overall effect (for example, because DSRIP and comparison hospitals made similar changes). To this end, we will examine whether cross-sectional variation in funding levels, areas of clinical focus, or certain types of patients (defined by clinical characteristics or recent hospital utilization) are associated with larger program effects.

We will do this by exploring two ways of estimating subgroup-specific effects: (1) by adding subgroups and subgroup-by-treatment interactions to the current model and (2) using an empirical Bayesian approach, which treats the subgroups as random effects. For instance, in the first approach, if we were interested in understanding whether DSRIP programs were more effective for individuals with behavioral health conditions, we would include an interaction term that includes time, whether the HSA is affected by the DSRIP, and whether the individual has a substance use disorder (creating a three-way, time-by-DSRIP-by-subgroup interaction). We would also interact other covariates with subgroup indicators as needed. Estimating these subgroup effects not only adds a more nuanced understanding of program impacts, it also serves as a robustness check on the comparative interrupted time series estimates.

H. Challenges and strategies for addressing them

The evaluation of the DSRIP demonstrations poses a number of challenges. First, the demonstration itself is complex because there are multiple levels of accountability and decision making, including federal, state, regional health organizations (in Texas), and provider organizations, which themselves often have multiple levels. Interventions are neither structured nor documented in a standard way. Thus it is less appropriate to think of this evaluation as assessing the effects of a single uniform initiative and more appropriate to think about it as a set of individual assessments of widely varying interventions supported by the DSRIP. We will seek to address this complexity by continually and carefully aligning our rapid cycle reports and our assessment of demonstration outcomes to ensure that we understand the demonstration as thoroughly as possible and that we incorporate this knowledge into the analyses of outcomes. Moreover, we have sought to create an evaluation that reflects the complexity of the demonstration by developing a conceptual and analytic framework that accommodates multiple levels, such as HSA and individual, but that also involves a small number of outcomes as well as evidence on the circumstances in which the DSRIP demonstrations are most effective.

Second, the demonstration is unfolding in the context of a rapidly changing health system, and many forces beyond the demonstration will affect the outcomes of interest. The pace of the change admittedly affects the precision with which the interrupted time series design can measure impacts. We plan to respond to this challenge by (1) using comparison groups that, as much as possible, are affected by these same forces; (2) supplementing the comparative interrupted time series estimates with cross-sectional estimates; and (3) designing models that incorporate a robust set of covariates to capture measurable changes in the environment. In addition, we will draw upon the rapid cycle reports to develop a qualitative understanding of what is driving demonstration impacts and where change likely results from other dynamics. We will use this knowledge to interpret results.

Third, we anticipate that the demonstration will not affect outcomes immediately. There are lags between program approval, project launch, and project impact. As a result, the interim evaluation focuses on the most immediate domains of delivery system transformation and clinical quality, rather than on longer-term outcomes such as readying providers for value-based purchasing, improving population health, and managing total cost of care.

Fourth, we expect the demonstration effect to be quite modest given the broad target population—all residents of the HSA, many of whom will be in good health. In response, we have chosen two measures (follow-up after ED visit measures) that target individuals who are likely to be in direct contact with participating institutions. However, we have balanced these outcomes with two other outcomes (ED visits and diabetes care) that are applicable to many residents in the HSA and independent of hospital contact, given the demonstration's broad focus. In addition, we will draw upon findings from the qualitative portion of the evaluation to identify the most relevant program features and community and patient characteristics for subgroup analyses.

Finally, as mentioned, the evaluation is limited by the available data, which influences several aspects of the design. First, the quality of the inpatient encounter data in our selected states limits the types of outcome measures we can construct. As a result, we selected measures that may serve as a signal of the broad changes in care that the demonstration aims to achieve and that can be measured with the available administrative data. When the available data pose additional limitations, we will modify the measures as needed. In the final evaluation, we will incorporate a broader range of outcome measures. In addition, because we are relying on Medicaid administrative data, we cannot estimate the impact of DSRIP programs on the uninsured, a key target population for these demonstrations. In the final evaluation, we plan to supplement the Medicaid administrative data with hospital discharge data to capture outcomes for uninsured populations.

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

References

Centers for Medicare & Medicaid Services, Center for Medicaid and CHIP Services. "Core Set of Children's Health Care Quality Measures for Medicaid and CHIP: Technical Specifications and Resource Manual for Federal Fiscal Year 2016 Reporting. Baltimore, MD: CMS, June 2016. Available at https://www.medicaid.gov/medicaid/quality-of-care/downloads/medicaid-and-chip-child-core-set-manual.pdf

Centers for Medicare & Medicaid Services, Center for Medicaid and CHIP Services. "Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid: Technical Specifications and Resource Manual for Federal Fiscal Year 2016 Reporting. Baltimore, MD: CMS, June 2016. Available at https://www.medicaid.gov/medicaid/quality-of-care/downloads/medicaid-adult-core-set-manual.pdf

Cook, Lawrence J., Stacey Knight, Edward P. Junkins, N. Clay Mann, J. Michael Dean, and Lenora M. Olson. "Repeat patients to the Emergency Department in a Statewide Database." *Academic Emergency Medicine*, vol. 11, no. 3, 2004, pp. 256–263.

Dartmouth Institute for Health Policy and Clinical Practice. "Geographic Crosswalks and Research Files." 2017. Available at http://www.dartmouthatlas.org/publications/reports.aspx. Accessed May 26, 2017.

Irvin, Carol V., Debra Lipson, Carey Appold, Maggie Colby, Katharine Bradley, Jessica Heeringa, Jenna Libersky, Vivian Byrd, and Julia Baller. "Medicaid 1115 Demonstration Evaluation Design Plan." Final report to the Centers for Medicare & Medicaid Services. Cambridge, MA: Mathematica Policy Research, May 2015. Available at https://www.medicaid.gov/medicaid-chip-program-information/by-topics/waivers/1115/downloads/evaluation-design.pdf.

Nysenbaum, Jessica Beth, Ellen Bouchery, and Rosalie Malsberger. "Availability and Usability of Behavioral Health Organization Encounter Data in MAX 2009." *Medicare & Medicaid Research Review*, vol. 4, no. 2. 2014.

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

APPENDIX TABLES

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

Table A.1. DSRIP program characteristics

Characteristic	CA[a]	TX[b]	MA[a]	NJ	NY
Approval date	11/1/2010	12/12/2011	12/20/2011	10/1/2012	4/14/2014
Expiration date	12/31/2020	12/31/2017	6/30/2017	6/30/2017	12/31/2019
Total program funding	$14.135B	$15.293B	$1.317B	$583M	$13.837B
Program funding per Medicaid beneficiary per month[c]	$14	$51	$10	$7	$35
Type of providers eligible to receive incentive payments	Designated public hospital systems and district/municipal public hospitals	Regional consortia of providers	Public and private acute hospitals with high Medicaid patient volume	Acute care hospitals	System of providers
Number of providers	55 hospitals	338 providers in 20 Regional Health Partnerships	7 hospitals	49 hospitals	91,603 providers in 25 performing provider systems
Broad or narrow eligibility	Medium All CA public hospital systems eligible	Broad A consortium in every region	Narrow 82 acute care hospitals in MA. Only 7 eligible.	Broad All hospitals eligible.	Broad Systems of thousands of providers
Number of projects	221[d]	1,450	47[d]	49	259

[a]Programs currently in renewal period

[b]Program in extension year

[c]To calculate the total program funding per Medicaid beneficiary per month, we used the total number of beneficiaries in the state as of 2014 from https://www.medicaid.gov/medicaid-chip-program-information/by-topics/data-and-systems/medicaid-managed-care/downloads/2014-medicaid-managed-care-enrollment-report.pdf.

[d]Number of projects in first waiver period

Table A.2. Project selection across DSRIP states and providers

Area of clinical focus[a]	California		Texas		Massachusetts		New Jersey	
	Project included on state menu	% of DPHs carrying out project	Project included on state menu	% of RHPs carrying out project	Project included on state menu	% of hospitals carrying out project	Project included on state menu	% of hospitals carrying out project
Primary care	X	100%	X	100%	X	86%	X	X
Physical and behavioral health integration	X	41%	X	90%	X	29%	X	2%
Appropriate care in appropriate settings	X	29%	X	100%	X	57%	X	22%
Diabetes care	X	X	X	X	X	29%	X	24%

[a]Additional clinical focus areas that were less common include access to care, behavioral health care, perinatal care, palliative care, nursing home care, dental care, disease or care management, medication management, patient safety, care transitions, health information technology, cardiovascular health, asthma, chronic renal failure, sexually transmitted infections, obesity, pneumonia, cognitive impairment, alternative payment models or value-based purchasing, and cost.