

The logo features the letters 'M50' in a bold, sans-serif font. The 'M' is orange with a red shadow, and the '50' is orange. A red horizontal line is positioned below the 'M50' text.

MATHEMATICA
Policy Research



Medicaid 1115
Demonstration
Evaluation
Design Plan

Delivery System Reform Incentive Payments

**Design Supplement:
Final Outcomes Evaluation
February 2019**

Julia Baller

Natalya Verbitsky Savitz

Jessica Heeringa

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

CONTENTS

A.	DSRIP DEMONSTRATIONS	1
B.	RESEARCH QUESTIONS AND OVERVIEW OF THE FINAL EVALUATION	2
C.	OUTCOME MEASURES AND DATA SOURCES	3
	1. Outcome measures	3
	2. Data sources.....	9
D.	PROPOSED EVALUATION DESIGN	11
	1. Difference-in-differences approach for Medicaid administrative data outcomes	11
	2. Simple interrupted time series approach for HCUP outcomes.....	17
E.	CHALLENGES AND LIMITATIONS.....	18
	REFERENCES.....	19
	APPENDIX A. DETAILED TABLES	21

TABLES

1.	Summary of DSRIP final outcome evaluation design	5
2.	DSRIP demonstration features and comparison groups, by state.....	11
3.	Data sources and potential list of covariates used to identify comparison HSAs	14
4.	Suggested subgroup characteristics.....	16
A.1.	Data availability, by state and data source	23
A.2.	DSRIP demonstration characteristics	25

FIGURES

1.	Logic model for DSRIP demonstrations.....	1
----	---	---

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

Delivery System Reform Incentive Payment (DSRIP) demonstrations, which are implemented under Medicaid section 1115 waiver authority, provide funding to health care providers to support infrastructure development, delivery system reform, and value-based payment (VBP), with the goals of improving quality of care and patient outcomes while reducing cost growth. The final evaluation, which we will conduct in 2020, will build on the methods used in the interim outcomes evaluation (Baller et al. 2018), conducted in 2016–2017. The final evaluation will provide the Centers for Medicare & Medicaid Services (CMS), state policymakers, and other stakeholders with information on (1) the effects of DSRIP demonstrations on delivery system transformation and clinical process measures, and (2) the circumstances under which DSRIP demonstrations have been more (or less) effective.

This document is a supplement to the Medicaid 1115 Demonstration Evaluation Design Plan (Irvin et al. 2015) and the DSRIP Design Supplement: Interim Outcomes Evaluation (Baller et al. 2017b) prepared by Mathematica Policy Research and submitted to CMS in May 2015 and June 2017, respectively.¹ This supplement presents our approach to the final outcomes evaluation of section 1115 DSRIP demonstrations.

The sections that follow describe the DSRIP demonstrations and the methods we intend to use to evaluate them, including the research questions, outcome measures and data sources, proposed evaluation design, and challenges and limitations to the evaluation.

A. DSRIP demonstrations

Since 2010, CMS has approved nine section 1115 DSRIP demonstrations (Center for Health Care Strategies 2016). The final outcomes evaluation will focus on earlier DSRIP demonstrations in California, New Jersey, New York, and Texas,² the four states with data available after demonstration implementation began and with sizable programs.³ Appendix Table A.1 lists all the states implementing DSRIP demonstrations and shows data availability for each state.

DSRIP demonstrations across states share a broad operational framework and many of the same goals (see Figure 1 for the DSRIP logic model). For example, to receive incentive payments, participating providers must implement projects that achieve specified milestones and metrics. However, they vary considerably in other respects across the study states (see Appendix Table A.2 for DSRIP demonstration characteristics by state). For example, in California and New Jersey, only hospital systems are eligible for incentive payments, whereas DSRIP programs in New York and Texas have more expansive provider eligibility criteria (for instance, Federally Qualified Health Centers are eligible to receive incentive funding). In addition to variation across

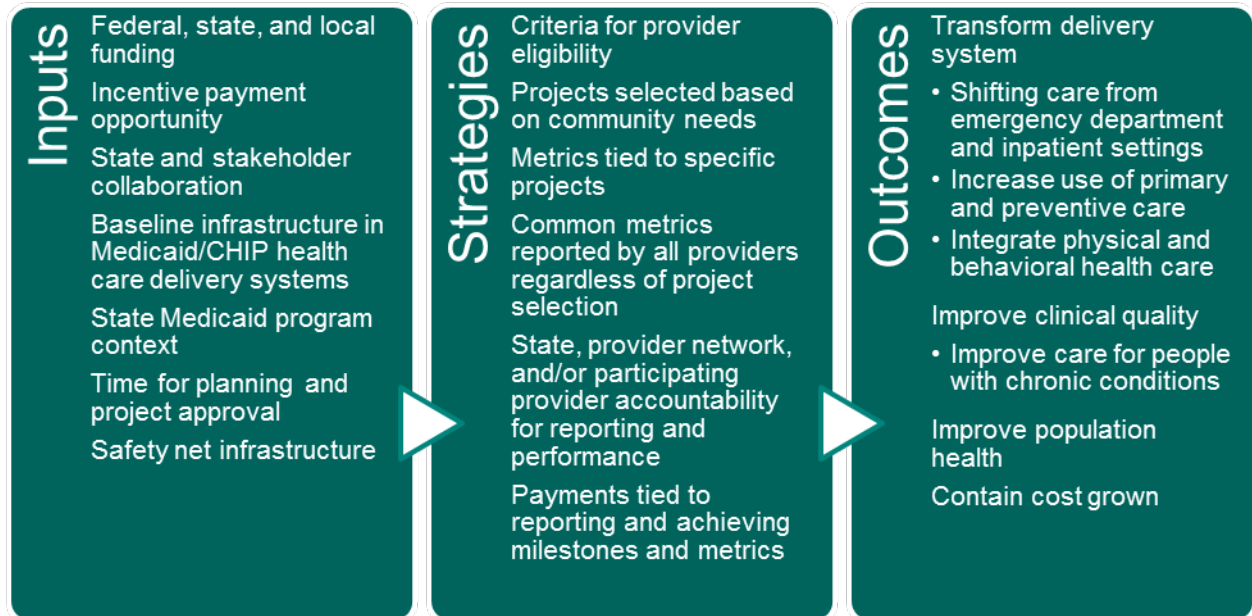
¹ There are several differences between the Medicaid 1115 Demonstration Evaluation Design Plan and this supplement. These differences are driven largely by data quality issues and feasibility concerns. For example, data quality limitations for cost-related data elements in the Medicaid encounter data preclude us from constructing cost-based measures.

² As of fall 2018, California, New Jersey, and Texas are implementing new DSRIP demonstrations. Because we will not have access to data from the current demonstration periods, the final outcomes evaluation will focus on understanding the impact of each state's initial DSRIP demonstration.

³ Per CMS guidance provided in 2015 during the initial evaluation design, we include states with DSRIP demonstrations with \$250 million or more in funding in the evaluation.

states, considerable variation exists across providers within a state in terms of the number and types of delivery system projects they implemented and the number and types of milestones and measures they reported. These sources of variation played a critical role in how we designed the final evaluation.

Figure 1. Logic model for DSRIP demonstrations



B. Research questions and overview of the final evaluation

The final evaluation will address two overarching research questions:

1. What is the overall effect of DSRIP demonstrations on key outcomes related to delivery system transformation and clinical processes in each state?
 - 1.1. What is the overall effect of DSRIP demonstrations on shifting care away from emergency department and inpatient settings?
 - 1.2. What is the overall effect of DSRIP demonstrations on use of primary care and preventive services?
 - 1.3. What is the overall effect of DSRIP demonstrations on use of behavioral health services?
 - 1.4. What is the overall effect of DSRIP demonstrations on clinical care processes?
2. Did DSRIP have differential effects on subgroups of interest?

The evaluation focuses on measures that are intended to capture a fundamental shift toward primary care and improved care coordination, which are intended to lead to declines in avoidable inpatient utilization use. Specifically, the evaluation will focus on 10 clinical process measures that (1) reflect the DSRIP demonstrations' overall goal of transforming the delivery system, as

characterized by an increased use in primary care and improved care coordination; and (2) are likely to respond relatively quickly to DSRIP projects.⁴ These outcomes include the following:

- Emergency department (ED) visits
- Preventable ED visits
- Prevention Quality Chronic Composite (Prevention Quality Indicator [PQI] 92)
- Adult ambulatory care visits
- Children and adolescents' access to primary care practitioners
- Tobacco use: screening and cessation intervention
- Behavioral health service use
- Antidepressant medication management
- Hemoglobin A1c testing
- Follow-up after ED discharge for patients with selected chronic medical conditions (such as asthma, chronic obstructive pulmonary disease [COPD], hypertension, or diabetes)

The outcomes will be drawn from two main data sources: (1) Medicaid administrative data and (2) the Agency for Healthcare Research and Quality's (AHRQ) Healthcare Cost and Utilization Project (HCUP) state inpatient databases. We describe the outcome measures and data sources in more detail in Section C.

The preferred analytic approach for the final evaluation will be a difference-in-differences model. In this approach, we will compare the changes in outcomes before and after each demonstration was implemented for people living in hospital service areas (HSAs) served by DSRIP providers with changes in outcomes for the same time period for people living in similar HSAs that are not served by DSRIP providers. When a suitable comparison group is not available, we will rely on a simple interrupted time series approach, which examines changes in both the level and trend of patient-level outcomes before and after each demonstration was implemented. We describe and summarize the design in more detail in Section D and Table 1.

C. Outcome measures and data sources

1. Outcome measures

We selected outcome measures for the final evaluation that closely align with the goals and activities of DSRIP demonstrations, are directly measureable using administrative data, and are most likely to be influenced by DSRIP incentives. The measures also reflect the priorities of CMS and the states for each DSRIP demonstration and include measures relevant to the most

⁴ In our 2015 evaluation design, we outlined research questions regarding the effects of DSRIP demonstrations on population health, per-capita costs, and VBP adoption. We are considering constructing population health measures, but data limitations preclude constructing a reliable cost measure. Furthermore, VBP progress was not a predominant focus of activity for most of the demonstrations in the final evaluation, and current administrative data do not reliably support construction of a measure of the proportion of Medicaid payment that is being made through VBP arrangements.

common clinical focus areas of the projects,⁵ endorsed measures, and measures in state DSRIP demonstration evaluations, when possible. In addition, our qualitative findings and reviews of state evaluations suggest that the demonstrations may not have observable impacts on health outcomes immediately (Baller et al. 2017a). Therefore, we focused on the most immediate domains of delivery system transformation and clinical processes, rather than on longer-term changes in health outcomes. In Table 1, we describe each outcome measure we intend to use to address each research question.⁶ Following Table 1, the outcomes for each research question are discussed in more detail.

⁵ As part of the interim outcomes evaluation, we developed a streamlined, comprehensive taxonomy of clinical focus areas that reflected the key goals of the DSRIP demonstrations to better understand state and provider clinical priorities. We mapped each project to one or more of the clinical focus areas. These areas included (but were not limited to) appropriate care in appropriate settings, primary care, behavioral health care, diabetes care, cardiac care, care transitions, and care coordination.

⁶ The final list of measures in the final report may be adjusted due to data quality, feasibility, and appropriate sample size.

Table 1. Summary of DSRIP final outcome evaluation design

Outcome measure used to address the research question	Measure description	Hypothesis	Sample or population subgroups to be compared ^a	Data sources	Analytic methods
Research question: What is the overall effect of DSRIP demonstrations on shifting care away from emergency department and inpatient settings?					
ED visits	Annual count of ED visits that did not result in an inpatient admission	ED visits that do not result in admissions will decline for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • Age • People with behavioral health conditions • People with select chronic conditions^d • Geographic characteristics 	Medicaid administrative data	Difference-in-differences
Preventable ED visits ^{b, c}	Annual count of preventable ED visits	Preventable ED visits will decline for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • Age • People with behavioral health conditions • Geographic characteristics 	Medicaid administrative data	Difference-in-differences
PQI 92: discharges for one of several chronic conditions	Quarterly count of discharges per 100,000 Medicaid beneficiaries and uninsured for: <ul style="list-style-type: none"> • Diabetes short-term complications • Diabetes long-term complications • COPD or asthma (in older adults) • Hypertension • Heart failure • Angina without procedure • Uncontrolled diabetes • Asthma in younger adults • Lower-extremity amputation among patients with diabetes 	Discharges for each of the chronic conditions included in the measure will decline after implementation of DSRIP.		HCUP data ACS data	Simple interrupted time series

Table 1 (continued)

Outcome measure used to address the research question	Measure description	Hypothesis	Sample or population subgroups to be compared ^a	Data sources	Analytic methods
Research question: What is the overall effect of DSRIP demonstrations on use of primary care and preventive services?					
Adult ambulatory care visits	Annual indicator of whether a beneficiary had an ambulatory care visit	Adult ambulatory care visits will increase for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • People with select chronic conditions • HSAs implementing primary care projects • Geographic characteristics 	Medicaid administrative data	Difference-in-differences
Children and adolescents' access to primary care practitioners	Annual indicator of whether a beneficiary has a visit with a primary care practitioner	Children and adolescents' access to visits with a primary care practitioner will increase for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • HSAs implementing primary care projects • Geographic characteristics 	Medicaid administrative data	Difference-in-differences
Tobacco use: screening and cessation intervention	Indicator of whether an adult beneficiary was screened for tobacco use within a year and received a tobacco cessation intervention if identified as a tobacco user	Tobacco screening and the provision of cessation intervention will increase for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs	<ul style="list-style-type: none"> • HSAs implementing behavioral health projects • Geographic characteristics 	Medicaid administrative data	Difference-in-differences
Research question: What is the overall effect of DSRIP demonstrations on use of behavioral health services?					
Behavioral health service use	Indicator of whether a beneficiary had a behavioral service visit	Outpatient behavioral health service use will increase for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • Age • HSAs implementing behavioral health projects • Geographic characteristics 	Medicaid administrative data	Difference-in-differences
Antidepressant medication management	Indicator of whether an adult beneficiary remained on antidepressant medication for at least (1) 84 days and (2) 180 days	Antidepressant medication management will increase for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • HSAs implementing behavioral health projects • Geographic characteristics 	Medicaid administrative data	Difference-in-differences

Table 1 (continued)

Outcome measure used to address the research question	Measure description	Hypothesis	Sample or population subgroups to be compared ^a	Data sources	Analytic methods
Research question: What is the overall effect of DSRIP demonstrations on clinical care processes?					
Hemoglobin A1c (HbA1c) testing (NQF 0059)	Indicator of whether an adult beneficiary with diabetes had an HbA1c test in a year	Hemoglobin A1c testing for adults with diabetes will increase for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • HSAs implementing diabetes projects 	Medicaid administrative data	Difference-in-differences
Follow-up within seven days after an ED visit for ACSC	Indicator of whether a beneficiary had an outpatient visit within seven days of discharge from the ED for ACSC	Follow-up within seven days after an ED visit for ACSC will increase for beneficiaries living in demonstration HSAs relative to those living in comparison HSAs.	<ul style="list-style-type: none"> • HSAs implementing primary care projects • Shortage areas • Geographic characteristics 	Medicaid administrative data	Difference-in-differences

^a We describe our planned subgroup analyses in Part D. The subgroup analyses will only be feasible for regressions relying on Medicaid administrative data. For those measures, we prioritized subgroups for which we hypothesize that DSRIP may have a differential impact on the outcome of interest.

^b For ED visits and preventable ED visits outcomes, a count of the visits is the preferred way of defining the outcome. However, if we find that these outcomes have nonstandard distributions (for example, bimodal), we will explore transforming them into binary indicators of whether a beneficiary had an ED or a preventable ED visit, respectively.

^c Preventable ED visits will be defined using New York University Emergency Department visit severity algorithm (Billings et al. 2000a, 2000b).

^d We will select targeted chronic conditions from the Chronic Condition Warehouse’s Medicaid Enrollee Supplemental File—Conditions Segment.

ACS = American Community Survey; ACSC = ambulatory care sensitive conditions (such as asthma, COPD, diabetes, and hypertension); DSRIP = Delivery System Reform Incentive Payment; ED = emergency department; HbA1c test = Hemoglobin A1c test routinely performed in people with type 1 or type 2 diabetes; HCUP = Healthcare Cost and Utilization Project; HSA = hospital service area.

1.1. What is the overall effect of DSRIP demonstrations shifting care away from emergency department and inpatient settings?

ED visits. We will count the total number of ED visits that do not result in an inpatient stay for each Medicaid beneficiary in a given year. We propose to apply the ED visits measure in the Core Set of children's Health Care Quality Measures for Medicaid and CHIP (Child Core Set), a measure commonly constructed for adults,⁷ to the beneficiaries ages 0-64 as a proxy that represents lack of access to primary care (Centers for Medicare & Medicaid Services 2018). If the DSRIP demonstrations increase access to primary care services, use of the ED should decline.

Preventable ED visits. Some ED visits are unavoidable, while other ED visits could have been prevented if the complaints were treated with a prompt visit to a primary care physician or a specialist (Bodenheimer et al. 2002). We will measure the total number of potentially preventable ED visits for each beneficiary in a given year. Following Billings et al. (2000a, 2000b), we will define preventable ED visits as ED visits for (1) conditions where immediate care was not required within 12 hours; (2) conditions that required treatment within 12 hours but that could have been diagnosed and treated in a typical primary care setting; and (3) conditions that required emergency care, but the emergency care could potentially have been avoided with the use of timely and effective primary care (such as flare-ups of asthma or diabetes). If the DSRIP demonstrations increase access to primary care services, the rate of preventable ED visits should decline.

PQI 92. AHRQ specifies PQIs to identify areas for which good outpatient care can potentially prevent the need for hospitalization. We will measure discharges per 100,000 Medicaid and uninsured individuals for one of the following conditions: diabetes with short-term complications, diabetes with long-term complications, uncontrolled diabetes without complications, diabetes with lower-extremity amputation, COPD, asthma, hypertension, heart failure, or angina without a cardiac procedure. This composite measure reflects the demonstration type's focus on transforming care and reducing avoidable hospital use. If the sample sizes are sufficient, we will also examine discharges for each individual measure that comprises the composite.

1.2. What is the overall effect of DSRIP demonstrations on use of primary care and preventive services?

Adult ambulatory care visits. Receiving appropriate ambulatory care, or care provided by health care professionals in outpatient settings, can reduce unnecessary inpatient and ED use. We will measure whether adults ages 18 and older have an ambulatory care visit during the year. This measure will help assess whether care is shifting away from unnecessary inpatient and ED use toward ambulatory care, which is a goal for all DSRIP programs.

Children and adolescents' access to primary care practitioners. Access to primary care is critical for the health of children and adolescents, and primary care has been found to reduce

⁷ California, New Jersey, New York and Texas include a measure of emergency department use in the adult population, according to Mathematica's analysis of metrics included in their DSRIP demonstrations.

preventable ED visits (Bloom 2011). We will measure whether children and adolescents ages 12 months to 18 years had a visit with a primary care practitioner.

Tobacco Use: Screening and Cessation Intervention. Cigarette smoking is one of the leading preventable causes of death in the US, and Medicaid beneficiaries are 68 percent more likely to smoke than the general population (Greene, Sacks, & McMenamain 2014). We will measure whether adult beneficiaries are screened for tobacco use and, if identified as a tobacco user, received tobacco cessation.

1.3. What is the overall effect of DSRIP demonstrations on use of behavioral health services?

Outpatient behavioral health service use. Despite the high prevalence of mental health disorders, substance use disorders, and co-occurring physical health conditions, unmet need for services persists across the United States (Han et al. 2017; Walker et al. 2015). DSRIP demonstrations seek to (1) improve access to behavioral health care, and (2) integrate physical and behavioral health services. To assess whether DSRIP is having the intended effects, we will measure whether adult beneficiaries had a behavioral health service visit within the year.

Antidepressant medication management. Major depression can result in severe impairment in daily functioning. Antidepressant medications can be an effective treatment approach, and clinical guidelines emphasize monitoring treatment effectiveness while managing side effects. We will measure whether adult Medicaid beneficiaries with major depression remain on antidepressants for an acute phase (at least 84 days) and a continuation phase (180 days).

1.4. What is the overall effect of DSRIP demonstrations on clinical care processes?

Comprehensive diabetes care: HbA1c testing. 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. We will measure HbA1c testing among those with diabetes to assess whether DSRIP demonstrations are influencing the delivery of diabetes care.

Follow-up after discharge from the ED for ambulatory care sensitive conditions (asthma, COPD, hypertension, and diabetes). Standards for high quality care indicate that many patients who visit the ED for these conditions should have a primary care visit soon afterward.⁸ More generally, people who do not receive follow-up care are more likely to be readmitted to the ED (Cook et al. 2004). 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.

2. Data sources

The evaluation will draw on two main data sources: (1) Medicaid administrative enrollment and claims data and (2) hospital discharge data from the HCUP state inpatient databases—to define the outcomes of interest (see Appendix Table A.1 for a full description of data

⁸ Based on discussions with Mathematica’s clinical experts.

availability). We will also draw on the American Community Survey (ACS) to construct the denominators for outcomes of interest relying on HCUP data.

Medicaid administrative data. To examine the impact of DSRIP demonstrations on Medicaid beneficiaries, we will use data derived from Medicaid enrollment files and claims paid to providers. We will have to combine several Medicaid administrative data sources because many states transitioned from the Medicaid Statistical Information System (MSIS) to a new, more standardized reporting format, T-MSIS, during our study period. For periods before a state's transition, we will use Medicaid Analytic eXtract (MAX), or the early version of MAX known as Alpha-MAX. For periods after a state's transition, we will use T-MSIS Analytic Files (TAF). MAX and Alpha-MAX are both research versions of state MSIS submissions; TAF is a research version of state T-MSIS submissions.⁹ These data will be available from 2009 through 2017 for California, New Jersey, New York, and Texas.¹⁰ TAF data are new, and a number of data quality and reliability questions are still outstanding. Data quality issues might affect the analyses we are able to conduct and the comparison states we are able to include.

Data limitations in each state influenced our selection of outcome measures, and they will likely impose some additional limitations on measure construction. For example, California and Texas do not have usable inpatient encounter records for adult beneficiaries at some points during our study period. As a result, we selected only measures that rely on outpatient data (including ED use), which will limit our ability to understand whether there is a shift in inappropriate inpatient use at the individual level.¹¹

HCUP state inpatient data. Another data source will be the HCUP data, the largest collection of longitudinal hospital care data in the United States. In contrast to Medicaid administrative data, HCUP data cover all people who receive care, including those without insurance—a key target population for DSRIP demonstrations. Because these data only include inpatient discharge records from hospitals, they cannot be used to study health outcomes for individuals. We will use HCUP data for New Jersey and Texas only, because we do not have access to HCUP data in other states that overlaps with their DSRIP demonstration periods.¹²

ACS. We will use the ACS to construct the denominator of the PQI chronic composite measure. The U.S. Census Bureau uses the annual ACS to collect social, economic, housing, and demographic indicators on the nation, states, counties, and local areas. We will use five-year estimates¹³ at the zip code level to obtain counts of Medicaid beneficiaries and the uninsured living within each HSA.

⁹ CMS develops MAX data as a more research-friendly version of MSIS files and TAF as a more research-friendly version of T-MSIS files.

¹⁰ TAF data will be reprocessed in 2019 to address known data quality issues for the period 2014-2017. We anticipate that the data files will be available for us by Fall 2019 and will be included in the analysis.

¹¹ Analyses using HCUP data will allow us to understand aggregate shifts in potentially preventable inpatient use.

¹² HCUP data are available only through 2011 in California and 2015 in New York. Therefore, we will not be conducting analyses using HCUP data in these states.

¹³ Depending on the population size of the geographic area of interest, the Census Bureau releases one-, three-, or five-year estimates based on ACS data. For this study, we will use five-year estimates, which are available for all areas reported by the Census Bureau, including zip codes.

D. Proposed evaluation design

The preferred analytic approach for the final evaluation will be a difference-in-differences model, the most rigorous feasible design. For Medicaid administrative data, we will compare pre- and post-implementation outcomes for people living in HSAs served by DSRIP providers with outcomes for people living in similar HSAs that are not served by DSRIP providers. For each demonstration state, the comparison group will be selected among similar HSAs in the same or neighboring region of the United States from states with high quality Medicaid administrative data (see Table 2).

When a suitable comparison group is not available or the data do not support the identification of a comparison group, as is true for analyses that rely on HCUP data, we will use a simple interrupted time series. This design will rely on a hospital as its own control by examining changes in both the level and the trend of hospital-level outcomes before and after the demonstration was implemented.

As mentioned earlier, states differ in both program design and data availability. To account for these differences, we will estimate separate regression models for each state and outcome combination.

Table 2. DSRIP demonstration features and comparison groups, by state

State	Demonstration approval date	DSRIP implementation start date	DSRIP eligible participants	Comparison group HSAs ^a
California	November 2010	November 2011	Designated public hospital systems and district/municipal public hospital systems ^b	CA
Texas	December 2011	October 2011	Regional health partnerships (statewide)	GA, KS, MO, NE, NM, UT
New Jersey	October 2012	January 2014	Acute care hospitals	CT, DC, DE, MA, OH, VA
New York	April 2014	April 2015	Performing provider systems (statewide)	CT, DC, DE, MA, OH, VA

^a An HSA is a collection of zip codes that define a hospital's catchment area. For analyses using Medicaid administrative data, we will construct similar demonstration and comparison groups by matching demonstration HSAs to similar HSAs within the states presented.

^b Eligibility for the California DSRIP demonstration, the subject of this evaluation, was restricted to designated public hospital systems. The successor program, the Public Hospital Redesign and Incentives in Medi-Cal program, extended eligibility to district/municipal public hospital systems.

1. Difference-in-differences approach for Medicaid administrative data outcomes

We will use a difference-in-differences design to estimate DSRIP demonstration impacts on the outcome measures listed in Section C. This approach examines whether outcomes in the demonstration group improved to a greater extent than outcomes in the comparison group after implementation of the DSRIP demonstration. Assuming that the demonstration and comparison groups are similar in the absence of the demonstration, this approach allows us to rule out alternate explanations for changes in outcome measures such as changes in policies or contextual factors that occur at the same time that DSRIP demonstration are implemented and affect the demonstration and comparison groups equally. An example is the implementation of key

provisions of the Patient Protection and Affordable Care Act in 2014, around the time DSRIP demonstrations were implemented in two of the four evaluation states.

Defining the demonstration group. We will define the population eligible for the demonstration as all continuously enrolled, full-benefit Medicaid beneficiaries up to age 64 who are not dually eligible for Medicare. The demonstration group will not be limited to beneficiaries who received care from providers participating in DSRIP for two key reasons. First, data limitations related to provider identifiers preclude us from accurately identifying beneficiaries who receive care from participating providers. Second, DSRIP demonstrations are intended to transform care for a broad group of Medicaid beneficiaries and the uninsured. The proposed approach will allow us to assess whether DSRIP demonstrations are effectively shifting care away from inappropriate hospitalizations and ED visits toward ambulatory settings.

In California and New Jersey, we will define the demonstration group as Medicaid beneficiaries residing within the catchment area of participating hospitals, and we will use the Dartmouth Atlas HSAs to define the hospital catchment areas (Dartmouth Institute for Health Policy and Clinical Practice 2017). In New York and Texas, where the DSRIP programs were implemented for the entire state, we will define the demonstration group as all Medicaid beneficiaries in the state.

For all four states, we will examine Medicaid beneficiary data from 2009 through 2017 calendar years. Depending on the state and its DSRIP demonstration implementation date, however, we will examine between three and seven years of data before implementation and between two and six years after demonstration implementation began.

Identifying the comparison group. Without random assignment, a key challenge to performing a credible difference-in-differences evaluation is to select a comparison group that is similar to the demonstration group on a range of key characteristics, but unaffected by the demonstration. Given 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 to identify a within-state comparison group. If this approach is not feasible, we will use an out-of-state comparison group from the same U.S. geographic region (defined by the US Census) as the demonstration state. Based on the work conducted in the interim outcomes evaluation, we plan to identify a within-state comparison group for California and out-of-state comparison groups for New Jersey, New York, and Texas.

To identify out-of-state comparison groups, we will be guided by two principles. First, the comparison group will be drawn from among non-demonstration states from the same or proximate geographic region as the demonstration state to ensure credibility of the comparison group. For New Jersey and New York, we will draw a comparison group from HSAs in the neighboring areas of the Northeastern and South Atlantic United States. For Texas, we will draw the comparison group from HSAs in neighboring areas of the Mountain, Southern, and Midwestern United States. Second, the states included in the comparison group pool need to

have good quality Medicaid administrative data.¹⁴ Table 2 shows the states that will form a potential comparison group pool for each demonstration state.

To identify the comparison group, we will match the HSAs in the demonstration and comparison states on a range of socioeconomic, health care access, and other key covariates measured before demonstration implementation (Table 3). We plan to draw on four data sources:

1. **Dartmouth Atlas of Health Care.** Access to and quality of health care varies greatly across the United States. In particular, a person's health care experience and quality of care depend, in large part, on the health care resources available in the area where the patient lives, including general care physicians, specialist care, hospitals, and mental and substance use disorder treatment facilities. The Dartmouth Atlas of Health Care provides data on hospital and physician capacity, as well as health care use at the national, regional, and local market levels based on Medicare data. We will use Dartmouth Atlas of Health Care data at the HSA¹⁵ level to identify a comparison group for each demonstration HSA in the evaluation.
2. **Health Resources and Services Administration (HRSA) data warehouse.** Some geographic areas of the United States have too few health care providers (for example, dental, mental and behavioral health, or primary care) to serve the needs of the local residents. HRSA works with states to identify these areas and to improve health care access in the specialty experiencing shortage by offering health care providers incentives to work in these geographic areas. A designation as a primary care or mental health care shortage area raises a possibility that programs, other than DSRIP, that are intended to increase access to physical and mental health providers and improve residents' health outcomes are being implemented in these areas. If not accounted for, differences on this covariate between the demonstration and comparison groups may result in biased estimates of the effects of the DSRIP program. We will use HRSA's designations of primary care and mental health care shortage areas to identify a comparison group for each demonstration state in the evaluation.
3. **2010 Census.** Access to care and health outcomes vary greatly by whether a person lives in an urban or rural area. Therefore, ensuring that demonstration and comparison HSAs are similar on this covariate is critical for a credible evaluation. To determine the urbanicity level of a given HSA, we will use the 2010 decennial census' urban-rural classification at the zip code level and Dartmouth Atlas crosswalk between zip codes and HSAs. Because HSAs could span both rural and urban areas, we plan to measure an HSA's urbanicity level by the percentage of the HSA's population who live in an urban area.
4. **ACS.** We will use the annual ACS (described above) to create key socioeconomic characteristics of interest for each HSA. We will use the estimates corresponding to the five years before the start of the DSRIP demonstration in each state.¹⁶ For example, because

¹⁴ Comparison states had to have usable encounter data from the MAX other therapy files from 2009-2012.

¹⁵ Dartmouth Atlas defines HSAs as "collection of zip codes whose residents receive most of their hospitalizations from the hospitals in that area." The United States is split into 3,436 HSAs.

¹⁶ The Census Bureau first added the health insurance coverage question to the ACS in 2008. Therefore, no estimates for health insurance coverage are available for 2007–2011, the five years before the start of the DSRIP demonstrations in California and Texas. Because the DSRIP demonstration is unlikely to affect health insurance

California and Texas began implementing DSRIP demonstrations toward the end of the 2011 calendar year, we will use 2007–2011 ACS estimates. Similarly, for New Jersey and New York, we will use 2009–2013 and 2010–2014 ACS estimates, respectively.

Table 3. Data sources and potential list of covariates used to identify comparison HSAs

Type of data (data source)	Possible covariates	Level
Health care access (Dartmouth Atlas)	Acute care hospital beds per 1,000 residents Discharges for ACSC per 1,000 Medicare enrollees	HSA level data
Health care shortage areas (HRSA data warehouse)	Primary care shortage areas Mental health shortage areas	Zip code level data aggregated to HSA
Locale (U.S. Census Bureau's 2010 Census data)	Urban/rural	Zip code level data aggregated to HSA
Socioeconomic data (U.S. Census Bureau's ACS) ^a	Population count Health insurance coverage ^b Poverty status Mean annual income Age Ethnicity Race	Zip code level data aggregated to HSA
Pre-demonstration health care outcomes (Medicaid data)	ED visits Adult ambulatory care visits	Individual-level data aggregated to HSA

^a For the ACS, we will use zip code level five-year estimates. Except for health insurance coverage for California and Texas, we will use estimates for the five years before implementation of the DSRIP demonstrations in a given state. In particular, for California, Texas, and their comparison groups, we will use 2007–2011 ACS estimates. For New Jersey and its comparison group, we will use 2009–2013 ACS estimates, and for New York and its comparison group, we will use 2010–2014 ACS estimates.

^b The U.S. Census first added the health insurance coverage question to the ACS in 2008. Therefore, for California, Texas, and their comparison groups, we will use 2008–2012 ACS estimates for this covariate.

ACS = American Community Survey; ACSC = ambulatory care sensitive conditions (such as asthma, COPD, diabetes, and hypertension); ED = emergency department; HRSA = Health Resources and Services Administration; HSA = hospital service area.

To match the two groups on these key characteristics, we will use a propensity score, which captures the probability that an HSA would have participated in the demonstration (if available) given the observed characteristics for matching. Matching on the estimated propensity score will allow us to efficiently balance the two groups on a set of covariates and to produce two groups that are similar on the *observed* characteristics without losing a large proportion of units (and thus limiting the generalizability of study findings) due to inability of finding an exact match (Rosenbaum & Rubin 1983). However, without random assignment, and to the extent that unobserved differences between the two groups may still exist, alternate explanations for study findings cannot be completely ruled out.

As discussed above, DSRIP implementation differs across the demonstration states. In particular, the timing of the implementation and the eligibility requirements for participation in DSRIP differ among demonstration states. Therefore, for each state, we will estimate a separate

coverage, we will use 2008–2012 estimates, the first time that these five-year estimates are available, for health insurance coverage for these two states.

propensity score model using the covariates measured in the five-year time period before implementation of the demonstration. After estimating the propensity scores for each HSA, we will match a demonstration HSA to one or more comparison group HSAs with similar propensity scores.

Once matched, we will confirm that the demonstration and comparison groups look similar (that is, mean differences between the two groups are smaller than 0.25 of a standard deviation for all key covariates). If we cannot achieve required balance for certain covariates, we will adjust for the differences between the two groups in our analytic approach by including these covariates as controls in the regression model.¹⁷

Estimating demonstration impact. To estimate the impact of the DSRIP demonstration in a given state, we will use a two-level regression equation, modeling an outcome (y) for person (i), in HSA (j) at time (t):

$$Y_{ij} = \beta_0 + \beta_1(Post_{ij}) + \beta_2(Demo_{ij}) + \beta_3(Post_{ij}) * (Demo_{ij}) + \beta_4(W_{ij}) + \alpha_j + \varepsilon_{ij}$$

This model includes three types of covariates:

$Post_{ij}$ is a post-demonstration period indicator, equal to 1 if the observation is in the post-period and equal to 0 if the observation is in the pre-period;

$Demo_{ij}$ is a demonstration indicator, equal to 1 if the HSA j is affected by the DSRIP demonstration and equal to 0 if HSA j is in the comparison group; and

W_{ij} are patient-level characteristics (such as age, gender, presence of chronic conditions) and characteristics of the patient's home zip code (such as mean income and whether it is classified as rural, a primary care shortage area, or a behavioral health shortage area).

The model also includes two error terms: an HSA-level random effect (α_j) and a residual error term (ε_{ij}). The hierarchical error structure of this model accounts for the nested nature of the data (patients are nested within HSAs). In other words, the model takes into account that people in the same HSA may be more similar than people in different HSAs.

The coefficient of interest, β_3 , represents the differential impact of the demonstration, relative to the comparison, on the *level* of the outcome variable.

Exploratory subgroup analyses. In addition to quantifying the overall effect of the DSRIP, it is important to understand whether DSRIP had differential effects on subgroups of interest. We plan to define the subgroups based on individual characteristics (for instance, age¹⁸ or presence

¹⁷ At times, regression models will include all propensity score matching covariates. Because the number of HSAs included in the regression models is relatively small, we will have to limit the number of covariates in these models.

¹⁸ Several measures in the final outcomes evaluation will be constructed for beneficiaries ages 0-64, but characteristics of the DSRIP programs may result in differential impacts on children versus adults. For instance,

of a behavioral health condition), particular dimensions of the DSRIP program, including project focus areas (primary care, diabetes care, and behavioral health care), and geographic characteristics¹⁹, so that these features may be refined in future demonstrations. Table 4 lists potential subgroups.

Table 4. Suggested subgroup characteristics

Individual characteristics	DSRIP demonstration characteristics	Geographic characteristics
<ul style="list-style-type: none"> • Age • Presence of behavioral health condition • Presence of selected chronic condition 	<ul style="list-style-type: none"> • HSAs implementing primary care projects • HSAs implementing behavioral health projects • HSAs implementing diabetes projects 	<ul style="list-style-type: none"> • Primary care shortage areas • Mental health shortage areas • Urban/rural

Note: We will finalize the subgroup analyses after we examine sample sizes carefully and determine whether we have sufficient statistical power to detect an impact.

We will estimate subgroup-specific effects in two ways. First, we will add subgroups and subgroup-by-demonstration interactions to the current model. We would also interact other covariates with subgroup indicators as needed. In particular, we will estimate the following regression model:

$$Y_{ij} = \beta_0 + \beta_1(Post_{ij}) + \beta_2(Demo_{ij}) + \beta_3(Post_{ij}) * (Demo_{ij}) + \beta_4(Subgroup_{ij}) + \beta_5(Post_{ij}) * (Subgroup_{ij}) + \beta_6(Demo_{ij}) * (Subgroup_{ij}) + \beta_7(Post_{ij}) * (Demo_{ij}) * (Subgroup_{ij}) + \beta_8(W_{ij}) + \alpha_j + \epsilon_{ij}, \text{ where}$$

Subgroup_{ij} is an indicator for the subgroup of interest.

Some analyses will require us to restrict the population included in the analysis, as opposed to including subgroup and subgroup-by-demonstration interactions. For example, to estimate the impact of DSRIP demonstrations on ED visits for adult beneficiaries, we will include only adults in our model.

All subgroup analysis will be considered exploratory. The study’s ability to examine different subgroups will depend on the size of these subgroups. Before we conduct subgroup analyses, we will explore the size of the subgroups and the implications on statistical power to detect an impact. Estimating these subgroup effects both adds a more nuanced understanding of program impacts and serves as a robustness check on the difference-in-differences estimates.

some providers may be receiving incentive payments for declines in ED visits for the adult population, which may result in differential impacts of DSRIP on adults as compared to children.

¹⁹ Geographic characteristics may play an important role in the effectiveness of DSRIP. For example, although demonstration programs in Texas and New York are statewide, each explicitly focused on encouraging provider participation, particularly in underserved areas. Understanding the extent to which the program influenced service use and clinical care processes in shortage areas may shed light on the effectiveness of this strategy.

2. Simple interrupted time series approach for HCUP outcomes

As mentioned earlier, we will examine HCUP data for New Jersey and Texas only—the two states where we have access to these data.²⁰ Because we do not have access to HCUP data for comparison states, analyses of HCUP outcomes will rely on a simple interrupted time series design. This design estimates 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.²¹

To estimate the impact of the DSRIP demonstration in a given state, we will use a two-level regression equation, modeling an outcome (y), in HSA (j), at time (t):

$$Y_{ij} = \beta_0 + \beta_1(\text{time}_{ij}) + \beta_2(\text{Post}_j) + \beta_3(\text{time}_{ij}) * (\text{Post}_j) + \beta_4(W_{ij}) + \alpha_j + \varepsilon_{ij}, \text{ where}$$

Time_{ij} measures time since the start of the DSRIP demonstration (in years);

Post_{ij} is an post-demonstration period indicator, equal to 1 if the observation is in the post-period and equal to 0 if the observation is in the pre-period; and

W_{ij} are characteristics of the HSA (such as mean income, whether it is classified as rural, primary care shortage area, or a behavioral health shortage area, percentage of residents uninsured or covered by Medicaid, hospital beds per resident, and number of hospitals), as well as quarter indicators to control for seasonal effects.

The model also includes two error terms: an HSA-level random effect (α_j) and a residual error term (ε_{ij}).

The coefficients of interest, β_2 and β_3 , represent the differential impact of the demonstration, relative to the comparison, on the *level* and the (linear) *trend* of the outcome variable. In particular, a statistically significant and substantively large β_2 would indicate that an immediate change in outcome occurred at the time of the DSRIP implementation, while a statistically significant and substantively large β_3 would indicate that the annual rate of change in the outcome variable is different in the pre- and post-demonstration periods.

We will examine the data and explore the feasibility of conducting subgroup analysis using HCUP data. If feasible, we will estimate subgroup-specific effects by adding subgroups and subgroup-by-demonstration interactions or restricting our analyses to the subgroups of interest.

²⁰ HCUP data are only available through 2011 in California and 2015 in New York, the years in which the demonstrations were implemented. This precludes us from using these data to estimate the effect of DSRIP in these states.

²¹ If after examining the data, we determine that an interrupted time series model does not fit the data well, we will discuss with CMS two alternative approaches: (1) conducting a pre-post analysis of HCUP outcomes data, and (2) excluding HCUP outcomes from the final evaluation.

E. Challenges and limitations

The evaluation of the DSRIP demonstrations poses several challenges. First, the demonstrations are complex because there are many levels of accountability and decision making, including federal, state, provider networks (in New York and Texas), and provider organizations, which themselves often have many levels. Interventions are neither structured nor documented in a standard way. We will seek to address this complexity by carefully drawing on our qualitative analyses of DSRIP implementation 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 each demonstration by developing a conceptual and analytic framework that accommodates many levels, such as HSA and individual beneficiary, as well as evidence on the circumstances in which the DSRIP demonstrations are most effective.

Second, DSRIP demonstrations unfold in a rapidly changing health system context, and many forces beyond the demonstration will affect the outcomes of interest. The pace of the change affects the study in two ways. First, it affects that study's ability to detect impacts, because changes that take longer to observe and are smaller are harder to detect without increasing the sample size. In the current study, increasing sample size is not feasible, because the number of affected HSAs is fixed in each state. Second, when evaluating an intervention in a highly dynamic environment, it is more challenging to distinguish the impacts of the intervention of interest from the impacts of contextual changes or changes in other programs or policies. We plan to respond to this challenge by (1) using comparison groups that, as much as possible, are affected by these same forces; and (2) designing models that incorporate a robust set of covariates to capture measurable changes in the environment. In addition, we will draw on 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, our ability to accurately estimate the impact of DSRIP demonstrations using Medicaid administrative data depends on our ability to identify a suitable comparison group. Our analyses assume that the comparison groups are a reasonably accurate estimate of what would have occurred in the absence of the demonstration. For analyses relying on Medicaid administrative data, we will use propensity score matching to identify an appropriate comparison group, and we will perform several diagnostic tests to ensure comparability across the demonstration and comparison groups. For analyses relying on HCUP data where a comparison group is not available, we will rely on trends in the outcomes of interest in the pre-period to predict what would have happened in the absence of DSRIP.

Finally, as mentioned, the evaluation is limited by the available data, and this influences several aspects of the design. Specifically, the quality of the Medicaid inpatient encounter data limits the types of outcome measures we can construct. Therefore, we selected measures that can be constructed with the available administrative data. In addition, the TAF has not yet been used for evaluation work. At this time, knowledge of the quality and completeness of these data is limited. We may discover additional limitations, making it necessary to modify the proposed measures and design.

References

- Agency for Healthcare Research and Quality. "Prevention Quality Chronic Composite: Technical Specifications." Version 2018. Rockville, MD: AHRQ, June 2018. Available at https://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2018/TechSpecs/PQI_92_Prevention_Quality_Chronic_Composite.pdf.
- Baller, Julia B., Jessica Heeringa, Carey Appold, Keanan Lane, Ryan Stringer, and James Woerheide. "Implementing Measurement Requirements in DSRIP Demonstration Programs." Final report to the Centers for Medicare & Medicaid Services. Washington, DC: Mathematica Policy Research, June 2017a. Available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/1115-ib7-508-dsrip-measurement-requirements.pdf>.
- Baller, Julia B., Marian V. Wrobel, Natalya Verbitsky-Savitz, Mariel Finucane, and Carol Irvin. "Delivery System Reform Incentive Payments: Medicaid 1115 Demonstration Evaluation Design Plan." Final report to the Centers for Medicare & Medicaid Services. Washington, DC: Mathematica Policy Research, June 2017b. Available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/eval-dsgn-dlvry-incntv-pymnts.pdf>.
- Baller, Julia B., James Woerheide, Keanan Lane, Natalya Verbitsky-Savitz, and Marian V. Wrobel. "Delivery System Reform Incentive Payments: Medicaid 1115 Demonstration Interim Evaluation Report." Final report to the Centers for Medicare & Medicaid Services. Washington, DC: Mathematica Policy Research, January 2018. Available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/dsrip-interim-eval-report.pdf>.
- Billings, John, Nina Parikh, and Tod Mijanovich. "Emergency Department Use: The New York Story." Issue brief. New York: Commonwealth Fund, November 2000a. Available at https://www.commonwealthfund.org/sites/default/files/documents/media_files_publications_issue_brief_2000_nov_emergency_room_use_the_new_york_story_billings_nystory.pdf.pdf.
- Billings, John, Nina Parikh, and Tod Mijanovich. "Emergency Department Use in New York City: A Substitute for Primary Care?" Issue brief. New York: Commonwealth Fund, November 2000b. Available at https://www.commonwealthfund.org/sites/default/files/documents/media_files_publications_issue_brief_2000_mar_emergency_department_use_in_new_york_city_a_substitute_for_primary_care_billings_eduse_433.pdf.pdf.
- Bloom, B., R.A. Cohen, and G. Freeman. "Summary Health Statistics for U.S. Children: National Health Interview Survey, 2011." *Vital Health Statistics*, series 10, no. 254. Hyattsville, MD: National Center for Health Statistics, 2012. Available at http://www.cdc.gov/nchs/data/series/sr_10/sr10_254.pdf.

- Bodenheimer, T., E.H. Wagner, and K. Grumbach. "Improving Primary Care for Patients with Chronic Illness: The Chronic Care Model, Part 2." *JAMA*, vol. 288, no. 15, 2002, pp. 1909–1914.
- Center for Health Care Strategies, Inc. "Delivery System Reform Incentive Payment: Mapping State Programs." Interactive map. Hamilton, NJ: CHCS, December 2016. Available at <https://www.chcs.org/resource/delivery-system-reform-incentive-payment-mapping-state-programs/>.
- Centers for Medicare & Medicaid Services. "Core Set of Children's Health Care Quality Measures for Medicaid and CHIP (Child Core Set): Technical Specifications and Resource Manual for Federal Fiscal Year 2018 Reporting." Baltimore, MD: CMS, Center for Medicaid and CHIP Services, February 2018. Available at
- 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." Lebanon, NH: Trustees of Dartmouth College, 2017. Available at <http://www.dartmouthatlas.org/publications/reports.aspx>. Accessed May 26, 2017.
- Greene, J., R.M. Sacks, and S.B. McMenamin. "The Impact of Tobacco Dependence Treatment Coverage and Copayments in Medicaid." *American Journal of Preventive Medicine*, vol. 46, no. 4, 2014, pp. 331–336.
- Han, B., W.M. Compton, C. Blanco, and L.J. Colpe. "Prevalence, Treatment, and Unmet Treatment Needs of U.S. Adults with Mental Health and Substance Use Disorders." *Health Affairs*, vol. 36, no. 10, 2017, pp. 1739–1747.
- 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>.
- Rosenbaum, Paul R., and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, vol. 70, no. 1, 1983, pp. 41–55.
- Walker, E.R., J.R. Cummings, J.M. Hockenberry, and B.G. Druss. "Insurance Status, Use of Mental Health Services, and Unmet Need for Mental Health Care in the United States." *Psychiatric Services*, vol. 66, no. 6, 2015, pp. 578–584.

APPENDIX A. DETAILED TABLES

This page has been left blank for double-sided copying

Table A.1. Data availability, by state and data source

State	Implementation start date	Total funding (millions)	MAX	Alpha-MAX	TAF ^a	Months of data in the post-period	HCUP	Include in final outcomes evaluation?
Demonstration states								
California	November 1, 2011 (DSRIP) July 1, 2016 (PRIME)	\$6,671 (DSRIP) \$7,464 (PRIME)	2009–2014	January 2015– September 2015	October 2015– December 2017	62	-	Yes
Kansas	January 1, 2013	\$60	2009–2012	January 2013– December 2013	January 2013– December 2017	48	-	No ^b
Massachusetts	July 2, 2012 (DSTI) July 1, 2017 (DSRIP)	\$13,192 (DSTI) \$1,800 (DSRIP)	2009–2013	January 2014– September 2014	October 2014– December 2017	0	-	No
New Hampshire	January 18, 2017	\$150	2009–2012	January 2013– December 2013	January 2014– December 2017	-	-	No
New Jersey	January 1, 2014	\$583	2009–2014	January 2015– September 2015	October 2015– December 2017	36	2009– 2015	Yes
New Mexico	January 1, 2015	\$29	2009–2012	January 2013– December 2013	January 2013– December 2017	24	-	No ^b
New York	April 1, 2015	\$12,837	2009–2013	-	January 2014 ^a – December 2017	18	-	Yes
Texas	October 1, 2011	\$26,118	2009–2012	January 2013– June 2016	July 2014– December 2017	63	2009– 2014	Yes
Washington	January 1, 2018	\$1,125	2009–2012	January 2013– December 2013	January 2014– December 2017	0	-	No
Comparison states								
Connecticut	N/A		2009–2013	January 2014– March 2015	April 2015– December 2017	N/A	-	Yes
DC	N/A		2009–2012	January 2013– December 2013	January 2014– December 2017	N/A	-	Yes
Delaware	N/A		2009–2012	January 2013– December 2013	January 2014– December 2017	N/A	-	Yes
Georgia	N/A		2009–2014	January 2015– September 2015	October 2015– December 2017	N/A	-	Yes
Kansas	N/A		2009–2012	January 2013– December 2013	January 2014– December 2017	N/A	-	Yes

Table A.1 (continued)

State	Implementation start date	Total funding (millions)	MAX	Alpha-MAX	TAF ^a	Months of data in the post-period	HCUP	Include in final outcomes evaluation?
Massachusetts	N/A		2009–2013	January 2014–September 2014	October 2014–December 2017	N/A	-	Yes
Missouri	N/A		2009–2014	January 2015–September 2015	October 2015–December 2017	N/A	-	Yes
Nebraska	N/A		2009–2012	January 2013–December 2013	January 2014–December 2017	N/A	-	Yes
New Mexico	N/A		2009–2012	January 2013–December 2013	January 2014–December 2017	N/A	-	Yes
Ohio	N/A		2009–2013	January 2014–September 2014	October 2014–December 2017	N/A	-	Yes
Utah	N/A		2009–2014	January 2015–September 2015	October 2015–December 2017	N/A	-	Yes
Virginia	N/A		2009–2012	January 2013–December 2013	January 2014–December 2017	N/A	-	Yes

^a As of December 2018, TAF data are available for 2015 and 2016. Data for 2014 and 2017 are currently under production and should be available by December 2018. Should the data be available by January 2019, they will be included in the analysis.

^b States will not be included because DSRIP funding is less than \$100 million.

DSRIP = Delivery System Reform Incentive Payment; DSTI = Delivery System Transformation Initiatives; HCUP = Healthcare Cost and Utilization Project; MAX = Medicaid Analytic eXtract; NA = Not Applicable; PRIME = Public Hospital Redesign and Incentives in Medi-Cal; TAF = T-MSIS Analytic Files.

Table A.2. DSRIP demonstration characteristics

Characteristic	CA ^a	TX ^b	NJ	NY
Approval date	11/1/2010	12/12/2011	10/1/2012	4/14/2014
Expiration date	12/31/2020	12/31/2017	6/30/2017	12/31/2019
Total program funding	\$14.135B	\$26.118B	\$583M	\$13.837B
Program funding per Medicaid beneficiary per month ^c	\$14	\$51	\$7	\$35
Type of providers eligible to receive incentive payments	Designated public hospital systems and district/municipal public hospitals	Regional consortia of providers	Acute care hospitals	System of providers
Number of providers	55 hospitals	338 providers in 20 Regional Health Partnerships	49 hospitals	91,603 providers in 25 performing provider systems
Broad or narrow eligibility	Medium All California public hospital systems eligible	Broad A consortium in every region	Broad All hospitals eligible	Broad Systems of thousands of providers
Number of projects	221 ^d	1,450	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.

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 ■
SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD



Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.