DEPARTMENT OF HEALTH & HUMAN SERVICES
Centers for Medicare & Medicaid Services
7500 Security Boulevard, Mail Stop S2-03-17
Baltimore, Maryland 21244-1850

State Demonstrations Group

CMS is pleased to release the Medicaid Section 1115 Demonstration Interim Evaluation Report on Medicaid Delivery System Reform Incentive Payments (DSRIP). This evaluation work was performed independently by Mathematic Policy Research. While the findings are preliminary and subject to data limitations typical of complex health services research, CMS believes that transparency will inform many stakeholders, including states and providers.

This interim evaluation presents baseline trends and preliminary findings about the overall effect of DSRIP demonstrations on key outcomes related to delivery system transformation and clinical quality for Medicaid beneficiaries. The analysis focuses on three states (California, New Jersey, and Texas), and three outcome measures that are intended to capture a fundamental shift toward primary care and improved care coordination:

(1) Emergency Department (ED) visits;
(2) Follow-up after an ED visit for patients with an ambulatory care sensitive condition (asthma, chronic obstructive pulmonary disease, hypertension, or diabetes); and
(3) Hemoglobin A1c testing for patients with diabetes.

Due to the difference in each state's DSRIP design and implementation, as well as the scope of the states' Medicaid programs, results are not compared between states. The findings vary across both states and measures, and in all three states there were varied findings associated with the demonstrations. The early results in the report demonstrate that there is more to be learned about effective Medicaid Delivery System Reform strategies, and more progress to be made in improving delivery systems for Medicaid beneficiaries. CMS will continue to monitor these demonstrations and other findings as they are available to help inform policy and to make adjustments to these types of demonstrations.

CMS is looking forward to receiving the final evaluation reports, which will focus on additional measures and will include additional years of data. CMS will release these reports when they are finalized likely in the Fall of 2019.

Sincerely,

Judith Cash
Director
State Demonstrations Group

Left Blank for Printing Purposes

Medicaid 1115 Demonstration Interim Evaluation Report

Delivery System Reform Incentive Payments

January 31, 2018

Julia B. Baller
James Woerheide
Keanan Lane
Natalya Verbitsky-Savitz
Marian V. Wrobel

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

ACKNOWLEDGEMENTS

We thank Lisa Shang, Travis Flemming, Alyssa Maccarone, Hoa Le, and Erick Geil for providing programming support. The programming team worked with multiple data sources, created analytical files, and produced descriptive tables and modeling output. Bob Schmitz, Priyanka Anand, and Mariel Finucane provided input into modeling approaches and did quality assurance reviews of the modeling specifications and all report chapters. Carey Appold and Maggie Colby also provided quality assurance reviews of all report chapters. Carol Irvin gave early input into the evaluation design. Bridget Gutierrez and Effie Metropoulos edited the report, and Stephanie Barna provided secretarial support.

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

CONTENTS

EXECUTIVE SUMMARY ix

I. INTRODUCTION 1

A. California 3

B. New Jersey 3

C. Texas 4

II. METHODS 7

A. Outcome measures 7

B. Study population 8

C. Comparison strategy 8

D. Data sources 9

E. Estimating demonstration effects 10

III. RESULTS 11

A. Descriptive statistics 11

B. Unadjusted trends and multivariate regression results 13

1. California 13

2. New Jersey 17

3. Texas 21

IV. DISCUSSION 26

A. What have other studies found? 27

B. Study limitations 29

C. Future research 30

REFERENCES 33

APPENDIX A: DSRIP PROGRAM CHARACTERISTICS

APPENDIX B: METHODS

APPENDIX C: REGRESSION RESULTS

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

TABLES

ES.1 Estimated impact of DSRIP: summary xi

I.1 Overview of design for the draft interim outcomes evaluation 3

III.1 Characteristics of study sample in baseline year 12

FIGURES

I.1 Delivery system reform incentive payment demonstration: logic model 2

III.1 Share of beneficiaries with an emergency department visit in California: demonstration and comparison groups by quarter, January 2009 through December 2014 14

III.2 Follow-up after an ED visit for an ACSC in California: demonstration and comparison groups by quarter, January 2009 through December 2014 15

III.3 Diabetes testing in California: demonstration and comparison groups by year, 2009–2014 16

III.4 Impact of DSRIP in California between 2009 and 2014: Change from pre to post DSRIP 17

III.5 Share of beneficiaries with an emergency department visit in New Jersey demonstration and New York comparison groups by quarter, January 2011 through December 2014 18

III.6 Follow-up after an ED visit for an ACSC in New Jersey demonstration and New York comparison groups by quarter, January 2011 through December 2014 19

III.7 Diabetes testing in New Jersey demonstration and New York comparison groups by year, 2011–2014 20

III.8 Impact of DSRIP in New Jersey between 2011 and 2014: Change from pre to post DSRIP 21

III.9 Share of beneficiaries with an emergency department visit in Texas by quarter, January 2009 through December 2013 22

III.10 Follow-up after an ED visit for an ACSC in Texas by quarter, January 2009 through December 2013 23

III.11 Diabetes testing in Texas by year, 2009–2013 24

III.12 Impact of DSRIP in Texas between 2009 and 2013: Adjusted patient outcomes pre and post DSRIP 25

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

EXECUTIVE SUMMARY

Delivery System Reform Incentive Payment (DSRIP) programs give federal funding to hospitals and other health care providers for projects designed to transform the delivery system and thereby improve quality of care and patient outcomes, reduce the cost of care, and prepare providers for value-based purchasing.

This interim outcomes evaluation examines the early DSRIP demonstration effects. The purpose of this evaluation is twofold: (1) to test the feasibility of the evaluation's analytic strategy (described in detail below), and (2) to present baseline trends and preliminary findings about the overall effect of DSRIP demonstrations on key outcomes related to delivery system transformation and clinical quality for Medicaid beneficiaries. The analysis is focused on DSRIP programs in California, New Jersey, and Texas, and on three outcome measures: (1) emergency department (ED) visits; (2) follow-up after an ED visit for patients with an ambulatory care sensitive condition (ACSC) (including asthma, chronic obstructive pulmonary disease, hypertension, or diabetes); and (3) hemoglobin A1C testing for patients with diabetes.

Given substantial differences between states, both in terms of their implementation of DSRIP and the design and scope of their Medicaid programs, we analyzed each state's data separately. We began by using unadjusted (raw) data to descriptively assess the trend in each outcome measure. We then performed multivariate regressions to estimate the relationship between DSRIP and each outcome measure after controlling for individual and community-level characteristics. In California, we relied on a difference-in-differences approach that compared outcomes before and after program implementation for Medicaid beneficiaries living in California communities affected by DSRIP to those living in similar California communities not affected by DSRIP. In New Jersey, we also relied on a difference-in-differences approach to compare outcomes for Medicaid beneficiaries living in communities affected by DSRIP in New Jersey to those living in similar communities unaffected by DSRIP in New York, before and after implementation of the program. In Texas, we relied on a simple interrupted time series approach in which we examined changes in both the level and the trend of patient-level outcomes before and after the demonstration was implemented. All analyses relied on data from the Medicaid Analytic eXtract, which means findings are limited to Medicaid beneficiaries, although DSRIP programs are also likely to impact the uninsured.

When we tested the feasibility of the analytic strategy, the results showed that our strategy for identifying a comparison group in California and New Jersey was essentially successful; the baseline trends in the outcomes of interest were usually similar for the demonstration and comparison groups. This pattern supports relying on a difference-in-differences technique to estimate the demonstration effect for the interim report.

Table 1 below reveals the main effect of DSRIP for each state on the outcomes of interest. These findings should be considered preliminary, and at this stage the interim outcomes evaluation may not identify an effect on the outcomes of interest even if the demonstration influences quality. Observable effects may take additional time to accrue because DSRIP programs initially focused on infrastructure development, which may not have had sufficient time to substantially impact the outcomes of interest during the period of our analysis.

The findings are inconsistent across both states and measures (Table 1). In California, based on a difference-in-differences model, DSRIP was associated with a smaller decrease in ED visits relative to the comparison group (a relatively worse outcome), no relative change in follow-up visits after an ED visit for a chronic condition, and a relatively smaller decrease in diabetes testing (a relatively better outcome). In New Jersey, the estimated effects were similar to those in California: DSRIP was associated with a smaller relative decrease in ED visits and no relative change in follow-up after ED visits, but an *increase* in diabetes testing for the demonstration group in the post-period, contrasting with a decrease in the comparison group (again, a relatively better outcome). In Texas, based on the interrupted time series model, DSRIP was associated with an increasing trend in ED visits in the post-period (a worse outcome), no effect on follow-up after ED, and mixed effects on diabetes testing.'

In the final outcomes report, scheduled for 2019, we plan to extend the analyses presented here with data from a longer post-intervention period. We also plan to include: (1) additional states that are implementing DSRIP programs; (2) additional comparison states that are not implementing DSRIP programs, potentially including a comparison state or states for Texas; (3) additional analyses that rely on comparative interrupted time series models instead of difference-in-differences or simple interrupted time series estimates; (4) analyses of the effect of DSRIP programs on additional measures of delivery system transformation, clinical quality, preparation for value-based payment, population health, and total cost of care; and (5) hospital discharge data to permit analyses that include uninsured individuals.

Table ES.1. Estimated impact of DSRIP: summary

		Outcome					
		ED visits[a]		Follow-up after an ED visit for an ACSC[b]		Diabetes testing [b]	
		Sign	Interpretation	Sign	Interpretation	Sign	Interpretation
				California			
1	Change in post-period—comparison group	−	Decrease	−	Decrease	−	Decrease
2	Relative change in post-period—DSRIP group	+	Smaller decrease than comparison (worse outcome)	.	Similar decrease to comparison	+	Smaller decrease than comparison (better outcome)
				New Jersey			
3	Change in post-period—comparison group	−	Decrease	.	No significant finding	−	Decrease
4	Relative change in post-period—DSRIP group	+	Smaller decrease than comparison (worse outcome)	.	No significant finding	+	Increase, despite decrease in comparison (better outcome)
				Texas			
5	Trend in pre-period	.	Increasing trend	.	No significant finding	−	Decreasing trend
6	Change in level post-period	.	No significant finding	.	No significant finding	−	Mixed effects in post-period (decrease in level, increase in trend relative to pre-period)
7	Change in trend in post-period	+	Increase in trend relative to pre period (worse outcome)	.	No significant finding	+	

Note: Each outcome was analyzed separately using a linear probability regression model that adjusted for sex, age, clinical conditions, median household income, number of beds per resident, and number of hospitals per HSA. In California and New Jersey, estimates are based on difference-in-differences models. Estimates in Rows 1 and 3 are based on the coefficient on an indicator for the post-period in a regression that controls for individual and community characteristics. Estimates in Rows 2 and 4 are based on the coefficient on an interaction of post-period*demonstration in a regression that controls for individual and community characteristics. In Texas, estimates are based on a simple interrupted time series model. Estimates in Rows 5–7 are based on coefficients on a time trend, an indicator for the post-period, and an interaction of post-period*trend in a regression that controls for individual and community characteristics. Only effects that are statistically significant at the 1 percent level are shown in this table.

[a]For ED visits, a decrease in rate suggests an improvement in performance.

[b]For follow-up visits and diabetes testing, an increase in rate suggests an improvement in performance.

ACSC = ambulatory care sensitive condition; ED = emergency department.

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

I. INTRODUCTION

Delivery System Reform Incentive Payment (DSRIP) programs, which are authorized as Medicaid Section 1115 demonstrations, provide federal funding to health care providers to conduct projects that aim to transform the delivery system and thereby improve quality of care and patient outcomes, reduce the cost of care, and prepare for value-based purchasing[1] (see Figure I.1 for the DSRIP logic model). This report presents findings from the interim evaluation, which focuses on demonstration programs in California, New Jersey, and Texas, the three states in which the program has been implemented for at least three years and that have at least one year of Medicaid data in the post-period.[2]

Although the DSRIP demonstrations share 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 like the one in California were primarily intended to provide financial support for safety net health systems that serve a high volume of Medicaid beneficiaries and the uninsured. Larger demonstrations in New York and Texas emphasize 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 in terms of the number and types of projects that are being implemented and the number and types of milestones and measures being reported. (See Appendix Table A.2 for details about the projects selected by providers).

Although DSRIP programs are intended to affect the way 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, but also for population health because high quality care should improve population health.

The purpose of the interim outcomes evaluation is twofold: (1) to test the feasibility of the analytic strategy (described in Chapter II); and (2) to present preliminary findings that address the following overarching research question:

- What was the overall effect of DSRIP demonstrations on key outcomes related to delivery system transformation and clinical quality?

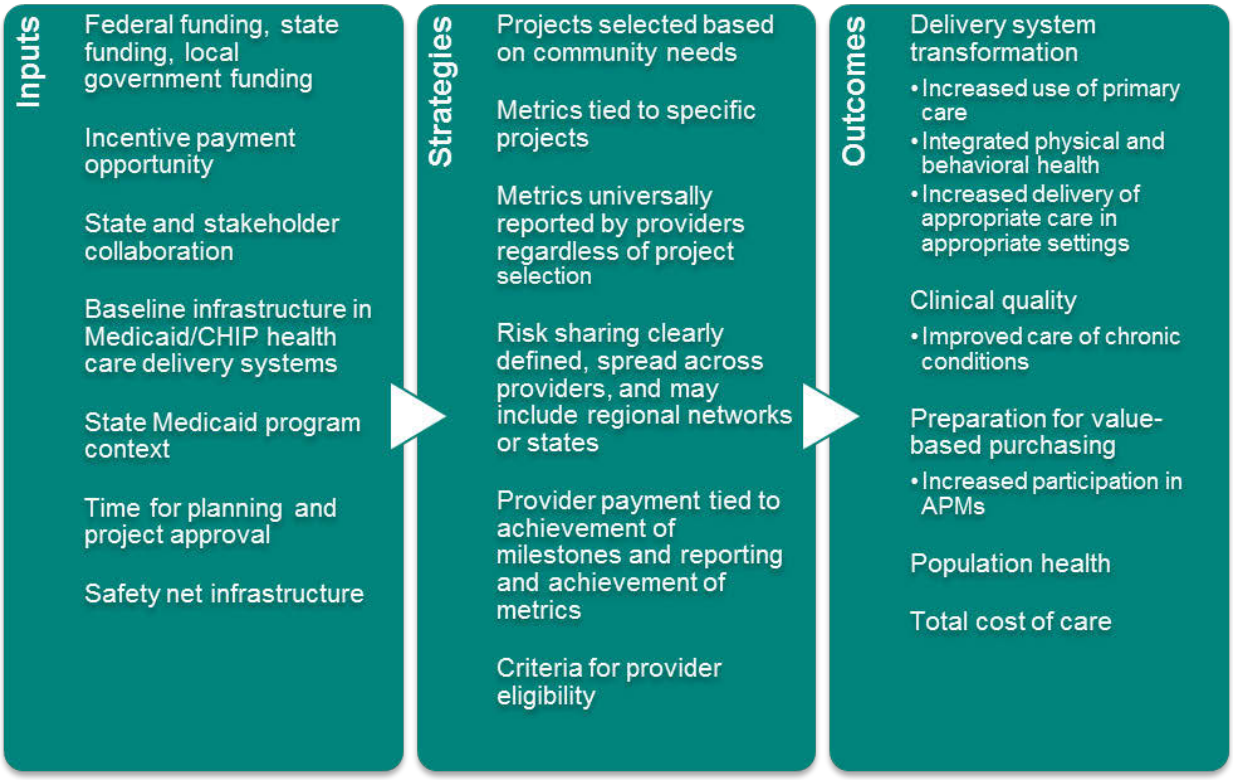
The analysis focuses on three clinical outcome measures that reflect the demonstration's overall purpose of transforming care and that are likely to respond relatively quickly to

[1] Preparing for value-based purchasing or adopting alternative payment models is an explicit goal of only some DSRIP programs, including California's Public Hospital Redesign and Incentives in Medi-Cal program, Massachusetts's Delivery System Transformation Initiatives and DSRIP programs, and New Hampshire's, New York's, and Washington's DSRIP programs.

[2] As of July 2017, seven states—California, Massachusetts, Texas, New Hampshire, New Jersey, New York, and Washington—have DSRIP programs. The final evaluation will include as many of these states as possible. However, our evaluation relies on claims data, which are delayed in certain states.

demonstration projects: (1) the rate of ED visits; (2) follow-up after an ED visit for an ambulatory care sensitive condition (ACSC) including asthma, chronic obstructive pulmonary disease [COPD], hypertension, or diabetes; and (3) hemoglobin A1c (HbA1c) testing for patients with diabetes. (See Appendix Table B.1 for additional measure details).

Figure I.1. Delivery system reform incentive payment demonstration: logic model.



The preferred analytic approach for the interim evaluation relies on a difference-in-differences technique, which estimates the average overall impact of DSRIP. In California and New Jersey, we compared patient-level outcomes for Medicaid beneficiaries living in hospital service areas (HSAs) served by DSRIP providers with outcomes for beneficiaries living in similar HSAs that were not served by DSRIP providers, before and after the demonstration was implemented. We used a within-state comparison in California, and an out-of-state comparison in New Jersey. When a suitable comparison group was not available, as was the case in Texas, we used a simple interrupted time series in which we examined changes in both the level and trend of patient-level outcomes before and after the demonstration was implemented.[3] All analyses relied on data from the Medicaid Analytic eXtract (MAX). (See Table I.1 for an overview of the interim outcomes evaluation design).

[3] We selected these models for the interim outcomes evaluation in part because they can be easily extended to comparative interrupted time series models, our anticipated method for the final outcomes evaluation in 2019.

Table I.1. Overview of design for the draft interim outcomes evaluation

Demonstration group	Comparison group	Pre–period	Post-period
California HSAs with participating DPHs	California HSAs without participating DPHs	2009–2011	2012–2014
New Jersey HSAs with participating acute care hospitals	New York HSAs	2011–2013	2014
Texas HSAs with participating providers	None	2009–2011	2012–2013

DPH = Designated public hospital system; HSA = hospital service area.

The remainder of this section describes programs in California, New Jersey, and Texas in detail. We then present methods (Chapter II), results (Chapter III), and conclude with a discussion and next steps (Chapter IV).

A. California

Shortly after the Patient Protection and Affordable Care Act (ACA) was signed into law in 2010, California began implementing its Section 1115 waiver, "Bridge to Reform." California's first-in-the-nation DSRIP program was designed to transform the state's safety net system in preparation for implementation of the ACA. More generally, the program sought to improve care for the state's Medicaid beneficiaries and the uninsured. Implementation of the state's DSRIP program began in November 2010 and concluded in October 2015.

Only designated public hospital systems (DPHs) were eligible for DSRIP, and all 21 DPHs in the state participated in the program. Over the course of the waiver, $3.4 billion was made available to participating DPHs for carrying out projects and meeting a combination of reporting and performance benchmarks. The projects and associated measures fell into five categories: (1) infrastructure development, (2) innovation and redesign, (3) population-focused improvement, (4) urgent improvement in care, and (5) HIV transmission (Centers for Medicare & Medicaid Services 2015). The total funding available to each DPH was determined based on the DPH system's cost, number of low-income individuals served, differences in quality infrastructure, and differences in patient populations. Although the DSRIP program in California aimed to make improvements in the five areas listed above, program funding was largely used for infrastructure investments.

California is now implementing its successor to the DSRIP program, the Public Hospital Redesign and Incentives in Medi-Cal program. This program was designed to build off of the foundational infrastructure improvements made during the original DSRIP program. It is more heavily focused on improving care delivery and incentivizing the adoption of risk-based alternative payment models (APMs) for managed care systems. The interim outcomes evaluation is only focusing on evaluating the first DSRIP program.

B. New Jersey

Approved in August 2013, the New Jersey DSRIP program aims to improve patient care for the state's low-income population by incentivizing reforms that improve access, enhance quality of care, and promote the health of patients and families (State of New Jersey Department of Health 2013). The DSRIP program replaced the state's Hospital Relief Subsidy Fund (HRSF), which provided payments to 55 acute care hospitals based on the amount of care delivered to

Medicaid beneficiaries and the uninsured. With total funding equaling $583 million, the DSRIP program provides no additional funding beyond what was previously available through the HRSF to the 49 hospitals that opted to participate. In spite of this, by tying this funding to implementing projects and reporting and improving upon measures, the Centers for Medicare & Medicaid Services (CMS) and the state expect improvements in clinical care and population health.

At the program's onset, each hospital selected a single project to implement over the course of the five-year demonstration. Each project focuses on one of the following chronic conditions: (1) asthma, (2) behavioral health, (3) cardiac care, (4) chemical addiction/substance use, (5) diabetes, (6) HIV/AIDS, (7) obesity, or (8) pneumonia. Participating hospitals most frequently selected projects focused on cardiac care (21 hospitals), diabetes (13 hospitals), and chemical addiction or substance abuse (5 hospitals).

Hospitals implement the projects over four stages: (1) infrastructure development, (2) chronic medical condition redesign and management, (3) quality improvements, and (4) population focused improvements. To receive funding, hospitals must carry out a specific set of activities within each implementation stage. These activities are associated with specific milestones. In addition, the hospitals must report a set of project-specific metrics (Stage 3 measures) and universal metrics (Stage 4 measures). Although most payments are tied to reporting Stage 3 and 4 measures, hospitals must also demonstrate improvement on a subset of Stage 3 measures to receive payments.

Although only acute care hospitals are eligible for DSRIP funding, several Stage 3 and 4 measures are focused on care delivered in outpatient settings. As a result, many participating hospitals are collaborating with community partners to implement the projects and assist with measure reporting.

The original demonstration was authorized in October 2012 for a five-year period. CMS granted an extension to continue the program through June 30, 2020.

C. Texas

The Texas DSRIP program, approved in December 2011, aims to improve quality of care, health status, patient experience, coordination and cost-effectiveness by transforming the way care is delivered to Medicaid beneficiaries and the uninsured. A total of $11.4 billion was made available to hospitals and other providers (Centers for Medicare & Medicaid Services 2015).

To achieve system transformation, Texas' program requires providers to organize into regional networks, known as a Regional Healthcare Partnerships (RHPs). Each RHP must be led by a public hospital or entity, which assumes responsibility for coordinating DSRIP activities. Within the RHP, participating providers (including, hospitals, counties, community health centers, and other provider types) select projects and report metrics that aim to address the specific needs of the communities in which they operate. There are 20 geographically defined RHPs with 338 total participating providers.

Each participating provider selects projects from four categories: (1) infrastructure development, (2) program innovation and redesign, (3) quality improvements, and (4) population-based improvements. A total of 1,450 projects are being implemented across the

state. The total value of each project is determined based on the project size and scope, as well as the provider's size and role in serving Medicaid and uninsured patients. Payments are earned primarily through pay-for-reporting, but some projects include pay-for-performance metrics in later years of the program. Similar to California, the incentive funding was used mostly for infrastructure development.

The original demonstration was approved for a five-year period from October 2011 through September 2016. CMS granted an initial extension to continue the program through December 2017 and a second extension to continue the program through September 30, 2021.

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

II. METHODS

In this chapter, we provide an overview of the outcome measures included in the evaluation, the population of beneficiaries included in our study sample, the way in which we selected our comparison groups, the main data sources, and the analytic approaches used to estimate the demonstration effects. Detailed technical descriptions are available in Appendix B, which contains additional details on how we constructed our outcome measures, our approach to identifying the study population, sample sizes across years for demonstration and comparison groups, our strategy for matching demonstration and comparison communities, Medicaid administrative claims data availability, and modeling specifications for our analyses.

A. Outcome measures

The analysis focuses on three outcome measures that are intended to capture a fundamental shift toward primary care and improved care coordination, which should lead to declines in avoidable hospital use. In selecting these measures, we sought to reflect CMS's and states' priorities for their DSRIP demonstrations, to include measures relevant to the most common clinical focus areas of the projects,[4] and to use endorsed measures and measures used by key project stakeholders. We also sought to analyze a small number of measures to make the effort more focused, ensure that the findings would be clear and easy to understand, avoid the loss of statistical power due to multiple comparisons, and design an analysis that was feasible given the time and resources available. As a result, we examined three clinical outcomes: (1) ED visits; (2) follow-up after an ED visit for patients with an ACSC (including asthma, COPD, hypertension, or diabetes); and (3) HbA1c testing for patients with diabetes.

ED visits. To our knowledge, no measures of availability or use of primary care among adults are currently endorsed.[5] As such, we applied the ED visits measure in the Medicaid Core Set of Child Quality Measures to the adult population. We rely on this measure as a proxy, where higher rates of ED visits suggest lack of access to primary care (Centers for Medicare & Medicaid Services 2017a). If the DSRIP demonstrations increase access to primary care services, use of the ED should decline.

Follow-up after ED visit for an ACSC. We examined follow-up after an ED visit for an ACSC as another measure of access to primary care and coordination across service settings. We measured the rate of follow-up within seven days of an ED visit for asthma, COPD, hypertension, and diabetes for those visits that did not result in an inpatient admission. Standards for high quality care indicate that many patients who visit the ED for these conditions should

[4] To better understand state and provider clinical priorities, we developed a taxonomy of clinical focus areas and mapped each project to one or more of the clinical focus areas. See Appendix Table A.2 for common areas of clinical focus and the extent to which each state adopted projects in these areas.

[5] To identify measures of availability and use of primary care among adults, we reviewed administrative measures endorsed by the National Quality Forum (NQF). Using this source, the team was not able to identify any 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.

have a primary care visit soon afterward.[6] More generally, individuals who do not receive follow-up care are more likely to be readmitted to the ED (Cook et al. 2004).

HbA1c testing. Finally, we measured HbA1c testing for beneficiaries with diabetes to assess whether DSRIP demonstrations were influencing the quality of diabetes care. Diabetes is a condition that is highly prevalent among Medicaid beneficiaries. In addition, DSRIP providers commonly select projects that focus on improving care for beneficiaries with diabetes. This measure is endorsed by the National Quality Forum (NQF) and is part of the Medicaid Core Set of Adult Health Care Quality Measures (Centers for Medicare & Medicaid Services 2017b).

B. Study population

The DSRIP demonstrations are intended to affect care for the entire community across a spectrum of providers. To reflect this, we defined the target population as all continuously enrolled, full-benefit Medicaid beneficiaries who reside within the catchment area of participating hospitals. We used the Dartmouth Atlas hospital service areas (HSAs) to define the hospital catchment areas (Dartmouth Institute for Health Policy and Clinical Practice 2017). We then defined the denominators and numerators for each measure within that population (illustrated in Appendix Figure B.1)**.** In keeping with the focus of the DSRIP programs and the definitions of our outcome measures, the sample consisted of adults, ages 18 to 64, who were not disabled.

C. Comparison strategy

The effect of the demonstration is 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). Given the differences in the states' DSRIP programs, we selected the analytic design and constructed comparison groups separately for each state, but used the same framework (see Appendix Figure B.2). The preferred analytic design was a difference-in-differences approach with an in-state comparison group, which avoids the challenge of state-to-state differences in policy factors such as what services or populations are covered by Medicaid and the extent of beneficiary cost-sharing. If this approach was not feasible, we used difference-in-differences approach with an out-of-state comparison group, followed by a simple interrupted time series with no comparison group.

In California, we compared the outcomes of interest for beneficiaries living in communities affected by DSRIP to the outcomes for beneficiaries living in similar communities that were not affected by DSRIP, before and after DSRIP implementation. Because participating hospitals in New Jersey cover nearly all HSAs in the state, we had to identify comparison communities outside the state. In New Jersey, we compared outcomes of interest for beneficiaries living in New Jersey communities affected by DSRIP to the outcomes of beneficiaries living in similar communities in New York, before and after DSRIP implementation. In Texas, we compared outcomes of interest after implementation of DSRIP to outcomes before DSRIP implementation

[6] This statement is based on discussions with Mathematica's clinical experts.

(See Appendix Figure B.3 to Appendix Figure B.5 for the number of Medicaid beneficiaries in each state's study sample.[7])

To create the comparison groups in California and New Jersey, we matched each demonstration community (defined as an HSA in which there was a provider participating in DSRIP) to one or more eligible comparison HSAs in which there were no providers participating in DSRIP. We used an exact matching strategy in which we chose four variables considered to be fundamental characteristics of an HSA: (1) the percentage of all discharges in the HSA that were for Medicaid beneficiaries; (2) the average number of beds in hospitals in each HSA; (3) whether the HSA was rural, urban, or a mix; and (4) whether the HSA had a high, medium, or low rate of discharges for Medicare beneficiaries with an ambulatory care sensitive condition. Each demonstration HSA could be matched to multiple comparison HSAs. (See Appendix Section B.1 for additional details about the hospital matching methods).

D. Data sources

To estimate the demonstration effect, we relied on several data sources, described below.

Medicaid enrollment and claims. Medicaid Analytic eXtract **(**MAX) data offer a comprehensive enrollment and claims history for each Medicaid beneficiary, which enable 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 for Texas. (See Appendix Table B.4 for details about MAX data availability and quality). MAX data are used to construct the three outcomes of interest, as well as the individual-level demographic and clinical characteristics.

American Hospital Association annual survey. We relied on the 2009 American Hospital Association (AHA) survey to define a variety of hospital-level characteristics (for example, the number of beds). The annual AHA survey provides a profile of over 6,400 hospitals in the United States. The survey includes facility-level variables on organizational structure, facilities and services, beds and utilization, staffing, expenses, physician arrangements, system affiliation, geographic indicators, accreditations and approval codes by credentialing organizations.[8]

American Community Survey. The American Community Survey (ACS), administered by the Census Bureau, collects annual demographic, housing, social, and economic data (US Census Bureau 2008). ACS data are used in this report for zip code-level household income and estimates of insurance coverage.[9] We used the five-year estimates for 2013 to determine the total

[7] In the final report, we will explore the possibility of creating a comparison group for Texas communities participating in the DSRIP program by using comparable communities from neighboring states, such as Arizona, Arkansas, Colorado, Kansas, Louisiana, New Mexico, and Oklahoma, that did not participate in the DSRIP program.

[8] For more on the AHA survey data collection methodology, see: https://www.ahadataviewer.com/about/data/.

[9] Zip Codes are represented as ZIP Code Tabulation Areas (ZCTAs) in Census Bureau surveys and are generalized representations of United States Postal Service (USPS) ZIP Code service areas. More information is available at https://www.census.gov/geo/reference/zctas.html.

population, median household income, and the share of the population either covered by Medicaid or are uninsured.

Health professional shortage areas data. The Health Resources and Services Administration designates certain regions, populations, or facilities as having shortages of primary care and mental health providers—known as health professional shortage areas (HPSAs). CMS publishes a list of all primary care and mental health HPSAs on an annual basis. We relied on the 2013 list to determine whether a beneficiary lives within a primary care or mental health HPSA.

E. Estimating demonstration effects

For each measure and state combination, we employed two main analysis methods. We first assessed the trend in each outcome measure descriptively by using unadjusted data. We then employed multivariate linear probability regression models to estimate the effect of DSRIP on the outcome measures of interest after controlling for beneficiary characteristics (for instance, age, gender, and clinical characteristics); zip-code level characteristics (for instance, median household income); and HSA-level characteristics (for instance, number of beds per resident and number of hospitals). In California and New Jersey, our main analyses relied on a difference-in-differences approach. In Texas, we relied on a simple interrupted time series approach. See Appendix B for more information on the analytic methods, including description of comparison groups and regression models used to estimate demonstration effects in Appendix Section B.2.

III. RESULTS

First, we examine descriptive statistics about the populations and communities that were included in the analyses (Table III.1). We then examine unadjusted trends and multivariate regression results for the outcomes of interest, separately for each state.

A. Descriptive statistics

In the baseline year (2011) in California, our study sample included 454,808 Medicaid beneficiaries in the demonstration HSAs and 488,251 in the comparison HSAs. [10] In both demonstration and comparison HSAs, about three-fourths of the beneficiaries (77 percent) were female, and almost two-thirds of the beneficiaries (64 percent) were between the ages of 18 and 34. The three most common chronic conditions were pulmonary, cardiac, and psychiatric conditions. Few beneficiaries lived in primary care or mental health care shortage areas. However, a higher percentage of beneficiaries in the comparison group lived in both types of shortage areas. In the demonstration group, most beneficiaries lived in HSAs with a high number of hospital beds per resident and with more than one hospital. In the comparison group, most beneficiaries lived in HSAs with a medium number of hospitals beds per resident, and close to half of beneficiaries lived in HSAs with one hospital. At baseline, outcome measures looked similar across demonstration and comparison HSAs. Six percent of beneficiaries had an ED visit in demonstration HSAs, compared to 7 percent in comparison HSAs. In both groups, about one in five beneficiaries who had an ED visit for an ACSC had a follow-up within seven days. About 6 in 10 beneficiaries in both groups who had diabetes had an HbA1c test.

At baseline (2013) for New Jersey, our sample included 89,137 Medicaid beneficiaries in New Jersey demonstration HSAs and 849,001 comparison beneficiaries in New York. In both demonstration and comparison HSAs, most beneficiaries were between the ages of 18 to 34. However, New Jersey had a considerably higher percentage of beneficiaries ages 18–25 (42 percent compared to 28 percent in New York) and a lower percentage of beneficiaries ages 55–64 (4 percent compared to 13 percent in New York). Roughly two-thirds of beneficiaries were female. The three most common conditions were pulmonary, cardiac, and psychiatric conditions. Few beneficiaries in New York and no beneficiaries in New Jersey lived in primary care or mental health care shortage areas. In New Jersey, most beneficiaries lived in HSAs with a high number of beds per resident and with one hospital. By contrast, beneficiaries living in comparison HSAs in New York lived in areas with a medium number of beds per resident and with more than one hospital. In the baseline period, a larger share of beneficiaries in New Jersey had an ED visit (19 percent in New Jersey versus 10 percent in New York); a smaller share of beneficiaries had a follow-up after an ED visit for an ACSC (22 percent in New Jersey versus 28 percent in New York); and a smaller share had diabetes testing (54 percent in New Jersey versus 65 percent in New York).

At baseline (2011) in Texas, our sample included 144,979 beneficiaries. Most (58 percent) beneficiaries were between the ages of 18 to 25, and most (82 percent) beneficiaries were female. The three most common conditions were pulmonary, psychiatric, and cardiac conditions. Sixteen percent of beneficiaries lived in a primary care shortage area, whereas 41 percent lived in

[10] Baseline is defined in California and Texas as 2011 and in New Jersey as 2013.

a mental health care shortage area. Most beneficiaries lived in areas with a low number of beds per resident and with more than one hospital. Twelve percent of beneficiaries had an ED visit; 25 percent of beneficiaries had a follow-up visit after an ED visit for an ACSC, and 69 percent of beneficiaries with diabetes had an HbA1c test.

Table III.1. Characteristics of study sample in baseline year

	California		New Jersey and New York		
	DSRIP HSAs % or mean	Comparison HSAs % or mean	NJ DSRIP HSAs % or mean	NY comparison HSAs % or mean	Texas % or mean
Total number of beneficiaries(N)	**454,808**	**488,251**	**89,137**	**849,001**	**144,979**
Individual characteristics					
Age					
18–25	37	37	42	28	58
26–34	27	28	24	23	26
35–44	21	21	18	19	13
45–54	12	12	13	17	3
55–64	3	2	4	13	1
Sex					
Female	77	77	70	61	82
Male	23	23	30	39	18
Chronic conditions					
Cardiac condition	14	15	27	33	11
Diabetes	5	6	8	11	4
Psychiatric condition	10	13	25	23	13
Substance use	4	5	17	15	4
Pulmonary condition	30	32	53	53	33
Zip code characteristics					
Primary care shortage area	1	3	0	2	16
Mental health care shortage area	4	8	0	7	41
Percent Medicaid and uninsured[a]					
Low	18	13	9	15	30
Medium	21	41	33	15	22
High	61	47	59	70	49
HSA characteristics					
Total HSAs (N)	**18**	**51**	**41**	**67**	**193**
Beds per resident[b]					
Low	5	32	20	34	47
Medium	41	55	29	57	42
High	54	13	51	9	11
Number of hospitals					
One	1	43	60	17	19
More than one	99	57	40	83	81
Outcome measures averaged over baseline year					
Share of beneficiaries with an ED visit	6	7	19	10	12
Follow-up after an ED visit for ambulatory care sensitive condition	21	22	22	28	25
Diabetes testing	57	61	54	65	69

Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from California from 2009 to 2014, from New Jersey and New York from 2011 to 2014, and from Texas from 2009 to 2013.

Table III.1 *(continued)*

Note: The baseline year is 2011 for California and Texas and 2013 for New Jersey and New York.

[a] For the purposes of the regression analyses, the percent Medicaid and uninsured was split into three categories—low, medium, and high. In California, the percent of Medicaid and uninsured in the low group was between zero and 27 percent, the medium group was between 28 and 33 percent, and the high group was between 34 and 54 percent. In New Jersey, the low group was between zero and 15 percent, the medium group was between 16 and 25 percent, and the high group was between 26 and 52 percent. In New York, the low group was between zero and 20 percent, the medium group was between 21 and 25 percent, and the high group was between 26 and 44 percent. In Texas, the low group was between zero and 33 percent, the medium group was between 34 and 38 percent, and the high group was between 39 and 65 percent.

[b] Beds per resident was split into three categories—low, medium, and high. The percentages of beneficiaries living in HSAs that fall within the low, medium, and high categories are presented. In California demonstration HSAs, the beds per resident in the low group was between zero and 0.001, the medium group was between 0.002 and 0.003, and the high group was between 0.004 and 0.019. In the California comparison HSAs, the beds per resident in the low group was between zero and 0.002, the medium group was between 0.002 and 0.003, and the high group was between 0.004 and 0.013. In New Jersey, the low group was between zero and 0.003, the medium group was between 0.003 and 0.004, and the high group was between 0.005 and 0.015. In New York, the low group was between zero and 0.003, the medium group was between 0.004 and 0.006, and the high group was between 0.007 and 0.025. In Texas, the low group was between 0.003, the medium group is between 0.004 and 0.006, and the high group is between 0.007 and 0.036.

ED = emergency department; HSA = hospital service area.

B. Unadjusted trends and multivariate regression results

1. California

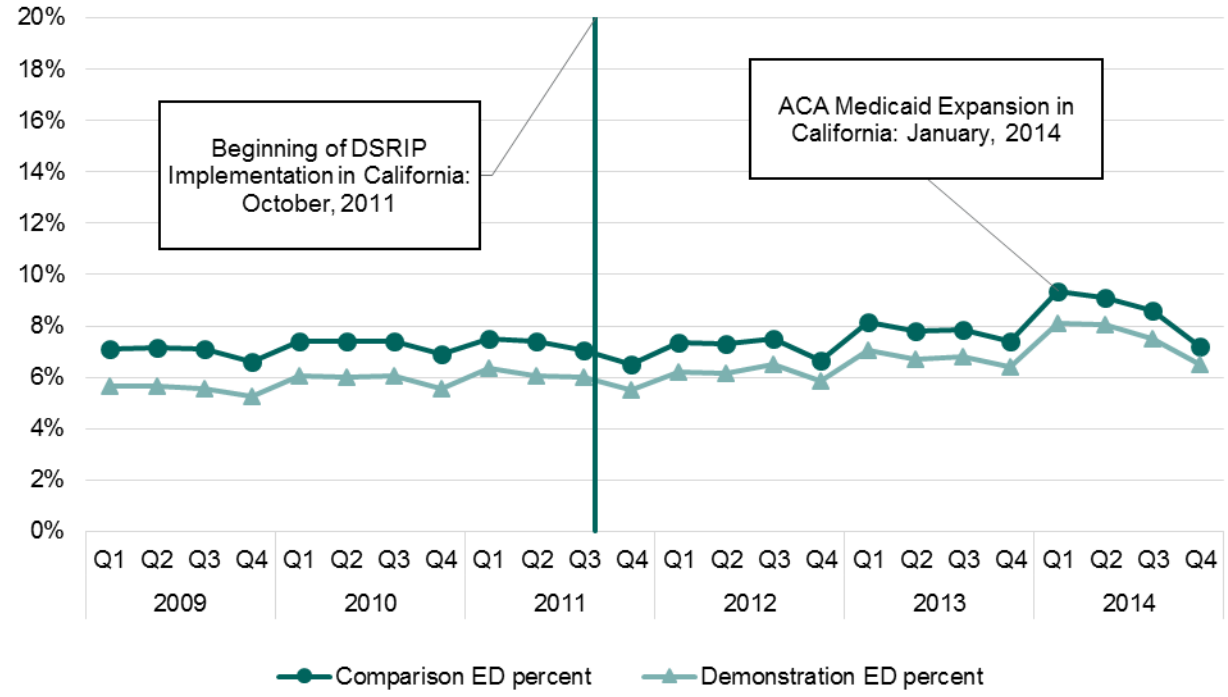
a. ED visits

The demonstration and comparison groups had similar pre-intervention trends in the share of beneficiaries with an ED visit (Figure III.1), which supports the use of an in-state comparison group and the difference-in-differences analytic approach. In addition, the pre-intervention trend highlights the presence of seasonal variation; rates are generally at their lowest in the fourth quarter of each year. In the unadjusted trends, the share of beneficiaries with an ED visit increased slightly for both groups after DSRIP was implemented, with a notable increase after Medicaid expansion in January 2014.

After we controlled for individual and community-level characteristics, we found that the probability of having an ED visit decreased in the demonstration group from the pre-period to the post-period (a favorable outcome). The probability of having an ED visit for the comparison group also decreased from the pre-period to the post-period.[11] However, the demonstration group experienced a smaller decrease than the comparison group, implying that the demonstration did not have an intended effect on the rate of the ED visits. (Figure III.4 and Table C.1).

[11] For all three outcome measures in California, there is a marked decline between the pre and post periods in the adjusted measure that exceeds the decline in the unadjusted measure. Measured rates of chronic conditions increased in the study population over time, which is likely one driver of this pattern.

Figure III.1. Share of beneficiaries with an emergency department visit in California: demonstration and comparison groups by quarter, January 2009 through December 2014



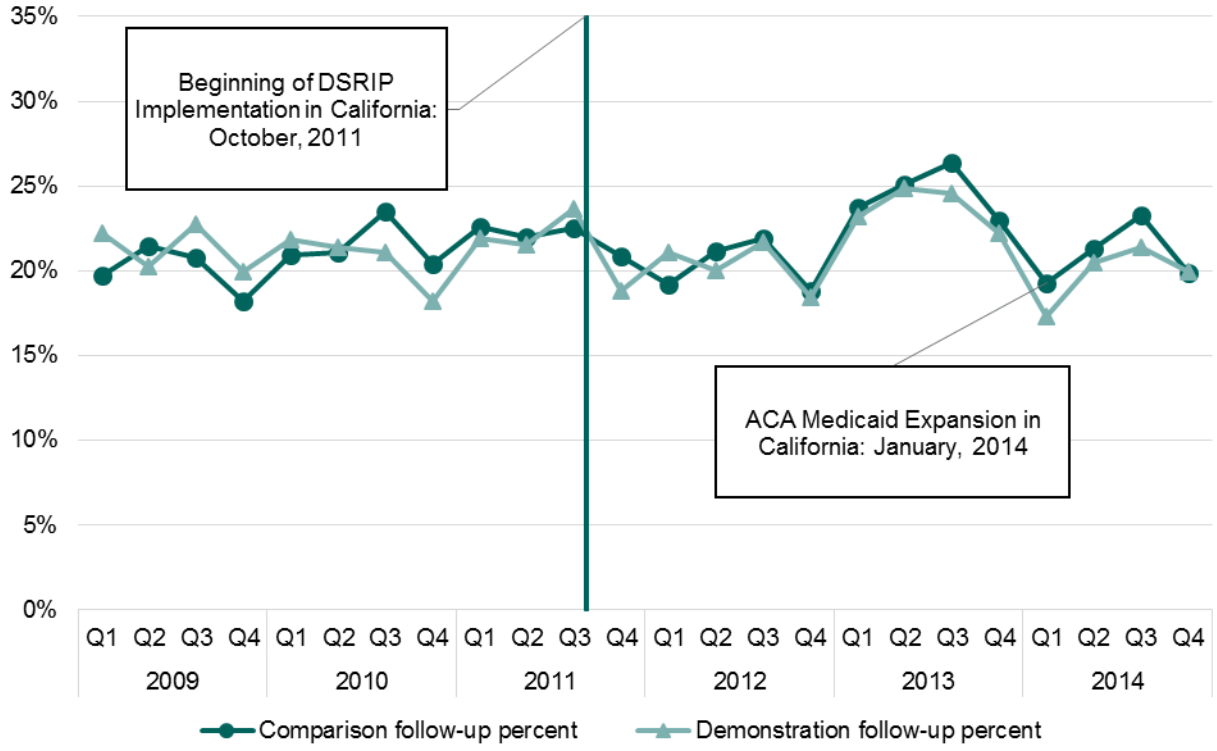
Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from California, 2009 to 2014.

b. Follow-up after an ED visit for an ACSC

The demonstration and comparison groups also had similar pre-period trajectories for the share of beneficiaries with a follow-up after an ED visit for an ACSC, which supports the use of an in-state comparison group and the difference-in-differences analytic approach. (Figure III.2). There was no consistent trend during the post-period. Although the rates for both groups were, on average, around 20 percent in the pre-period, they increased to roughly 22 percent for both groups in the post-period.

After controlling for individual and community-level characteristics, the probability of having a follow-up after an ED visit for ACSC decreased from the pre-period to the post-period in the demonstration group (an unfavorable outcome). For the comparison group, the probability also decreased in the post-period compared to the pre-period. Since the decrease for the demonstration and comparison groups were not statistically different from each other, we conclude that the demonstration did not have an effect (Figure III.4 and Table C.1).

Figure III.2. Follow-up after an ED visit for an ACSC in California: demonstration and comparison groups by quarter, January 2009 through December 2014



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from California, 2009 to 2014.

c. Diabetes testing

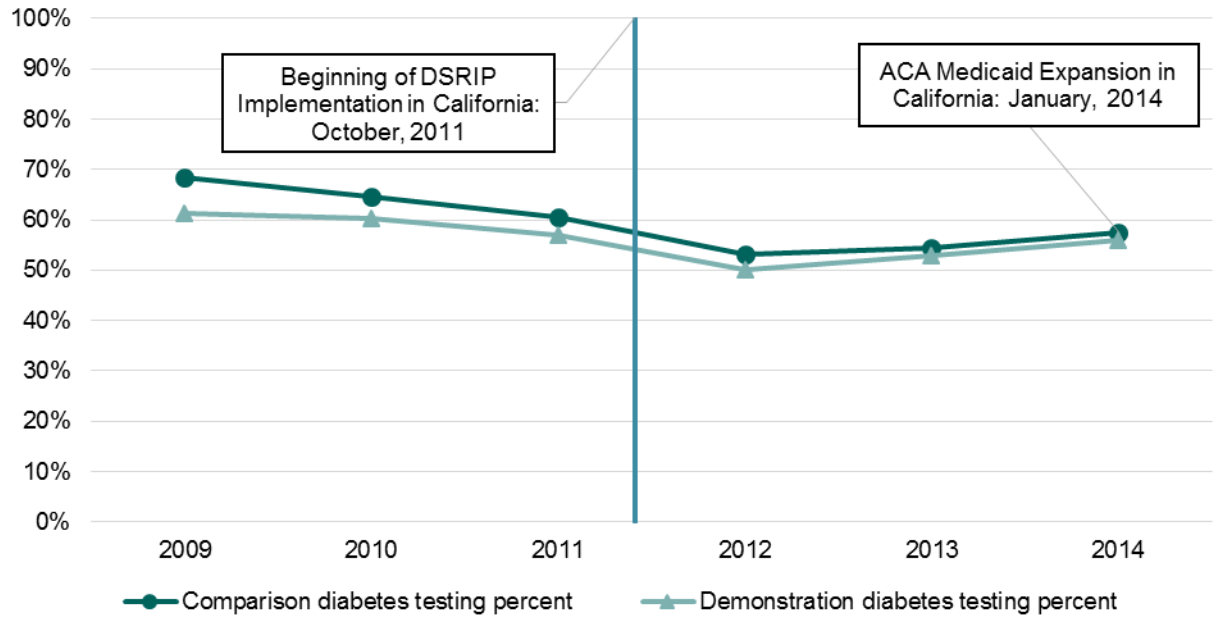
Finally, demonstration and comparison groups had similar trends in the baseline period for the percentage of beneficiaries with diabetes who had an HbA1c test during the year, which supports the use of an in-state comparison group and the difference-in-differences analytic approach. (Figure III.3). Both groups saw a decrease in the share of beneficiaries with an HbA1c test over the pre-period and a dip right after DSRIP was implemented, followed by an increase in the post-period.[12]

After we controlled for individual and community-level characteristics, we found that the probability of having a diabetes test decreased in the demonstration group from the pre-period to the post-period (an unfavorable result). However, in the comparison group, the adjusted probability of having a diabetes test decreased by a larger percentage in the post-period

[12]In light of the potential for a secular trend (for example, the seeming declines in diabetes testing in California) and the potential for this trend to differ between the demonstration and comparison groups, we plan to explore the use of comparative interrupted time series models for the final report. We did not have enough data to use these models in the interim report.

compared to the pre-period.[13] This suggests that the demonstration had a positive effect on diabetes testing. (Figure III.4 and Table C.1).

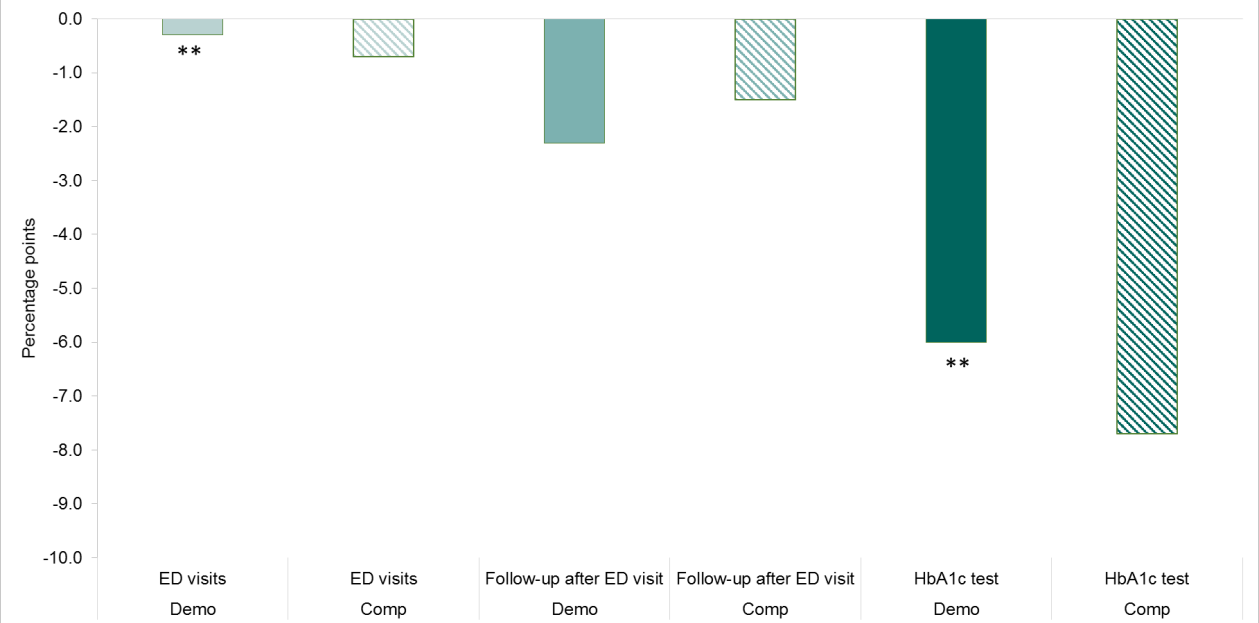
Figure III.3. Diabetes testing in California: demonstration and comparison groups by year, 2009–2014



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from California, 2009 to 2014.

[13] This finding is likely related to the rising rates of chronic conditions noted earlier. The rate of diabetes testing likely rises as more people have chronic condition. After controlling for chronic conditions, the adjusted rate of testing is negative.

Figure III.4. Impact of DSRIP in California between 2009 and 2014: Change from pre to post DSRIP



Source: Mathematica analysis of MAX and Alpha-MAX data from California, 2009 to 2014.

Note: This figure presents findings based on a difference-in-differences analysis for three outcomes: emergency department [ED] visits, follow-up after ED visit for an ambulatory care sensitive coditions, and diabetes (HbA1c) testing. Each outcome was analyzed separately using a linear probability regression that adjusted for sex, age, clinical conditions, median household income, number of beds per resident, and number of hospitals per HSA. Each bar represents a difference between one of the adjusted outcomes before and after DSRIP for one of two groups: demonstration (or DSRIP) HSAs and similar comparison (non-DSRIP) HSAs. ED visits and follow-up after an ED visit for ambulatory care sensitive conditions are measured at the person-quarter level. Diabetes testing is measured at the person-year level. * $p < 0.01$, ** $p < 0.001$.

2. New Jersey

a. ED visits

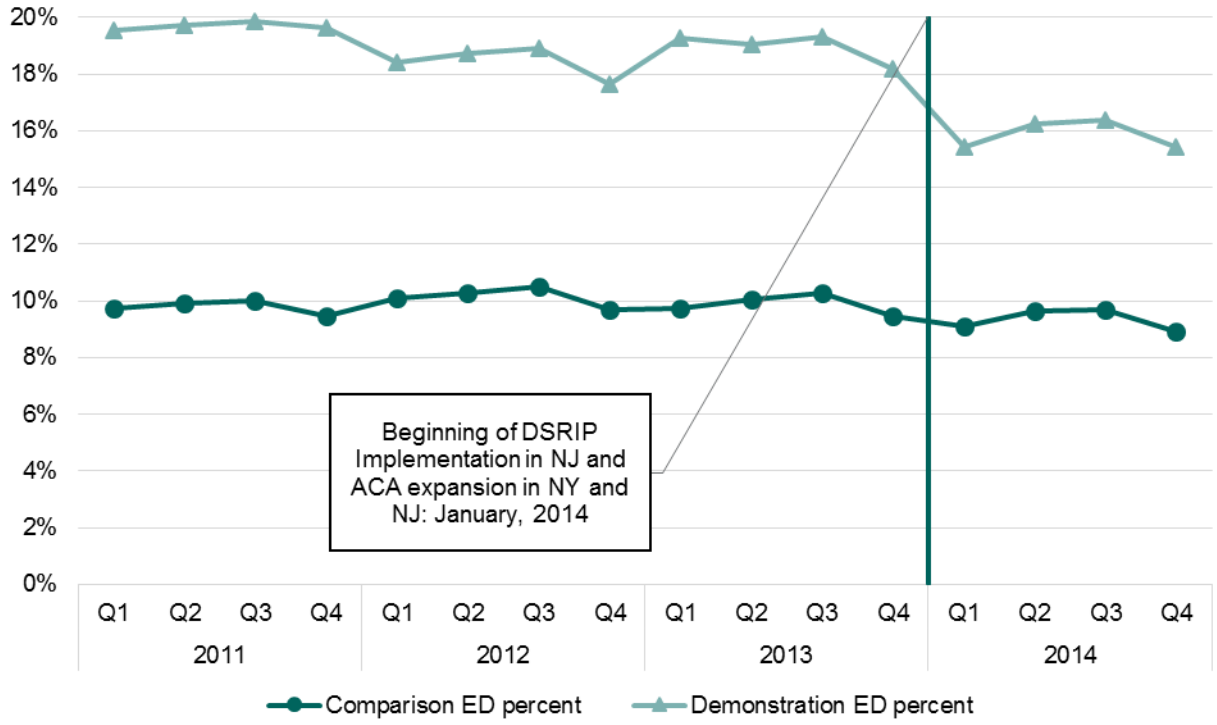
Trends in the share of beneficiaries with ED visits were similar across the demonstration and comparison groups in the pre-period (Figure III.5). The rate was considerably higher in New Jersey than in New York, but both states had relatively flat trajectories with some seasonal variation in the pre-intervention period. After a dip in ED visits rates around the time of DSRIP implementation and Medicaid expansion, which was particularly pronounced in New Jersey, the rates seemed to level off in the following three quarters.[14]

After controlling for individual and community-level characteristics, we found that the probability of having an ED visit decreased in the demonstration group from the pre-period to the post-period (an unfavorable outcome). In the comparison group, the probability of having an ED visit also decreased in the post-period compared to the pre-period. However, the probability of having an ED visit decreased less in the demonstration HSAs than in the comparison HSAs,

[14] Given these differences, in future work, we plan to examine impacts for groups of beneficiaries who were relatively unaffected by the Medicaid expansions in order to understand the sensitivity of our results to this major programmatic change. We also plan to assess the possibility of a multi-state comparison group for New Jersey.

suggesting that the demonstration did not have an intended effect on the rate of ED visits (Figure III.8 and Table C.2).

Figure III.5. Share of beneficiaries with an emergency department visit in New Jersey demonstration and New York comparison groups by quarter, January 2011 through December 2014



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from New Jersey and New York, 2011 to 2014.

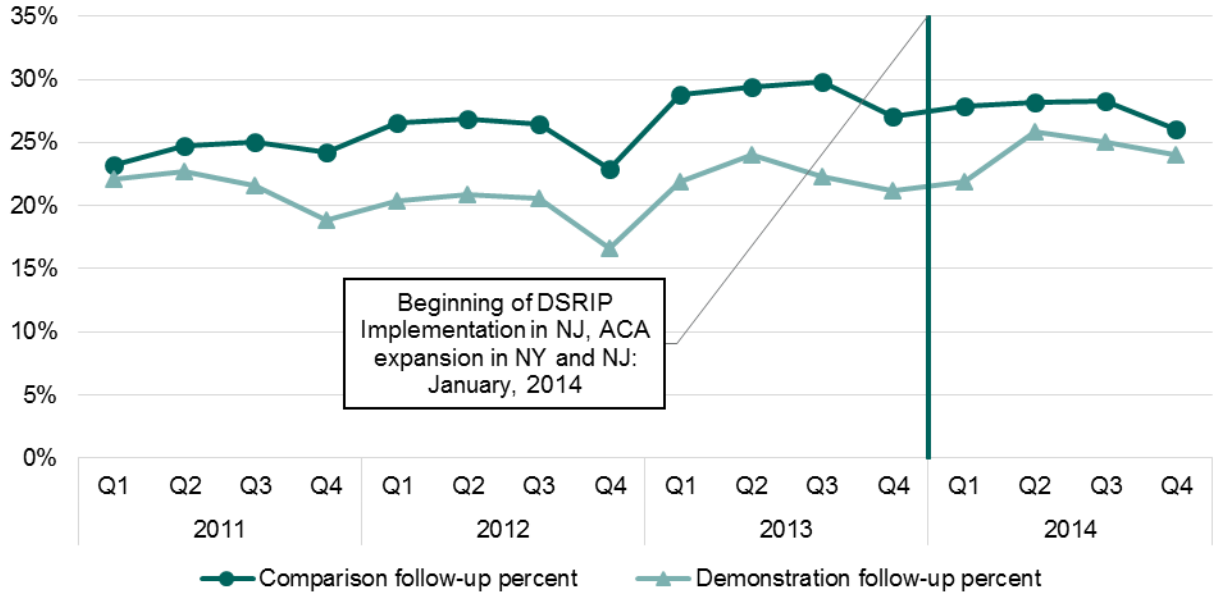
Note: text says that ED dip was particularly pronounced in New York, however the graph shows that the biggest dip occurred for the demonstration ED percent.

b. Follow-up after an ED visit for an ACSC

Figure III.6 shows the share of beneficiaries with a follow-up visit after an ED visit for an ACSC. The trends were similar across demonstration and comparison groups. In the pre-period, follow-up visits were more likely over time in both groups, with seasonal variation. In the post-period, rates stayed relatively level in the comparison group but they increased starting in the second quarter for the demonstration group.

After controlling for individual and community-level characteristics, the probability of having a follow-up after an ED visit for ACSC increased in the demonstration group from the pre-period to the post-period. While the probability decreased in the comparison group during the same time, the difference between the two groups was not statistically significant, suggesting no demonstration effect (Figure III.8 and Table C.2).

Figure III.6. Follow-up after an ED visit for an ACSC in New Jersey demonstration and New York comparison groups by quarter, January 2011 through December 2014



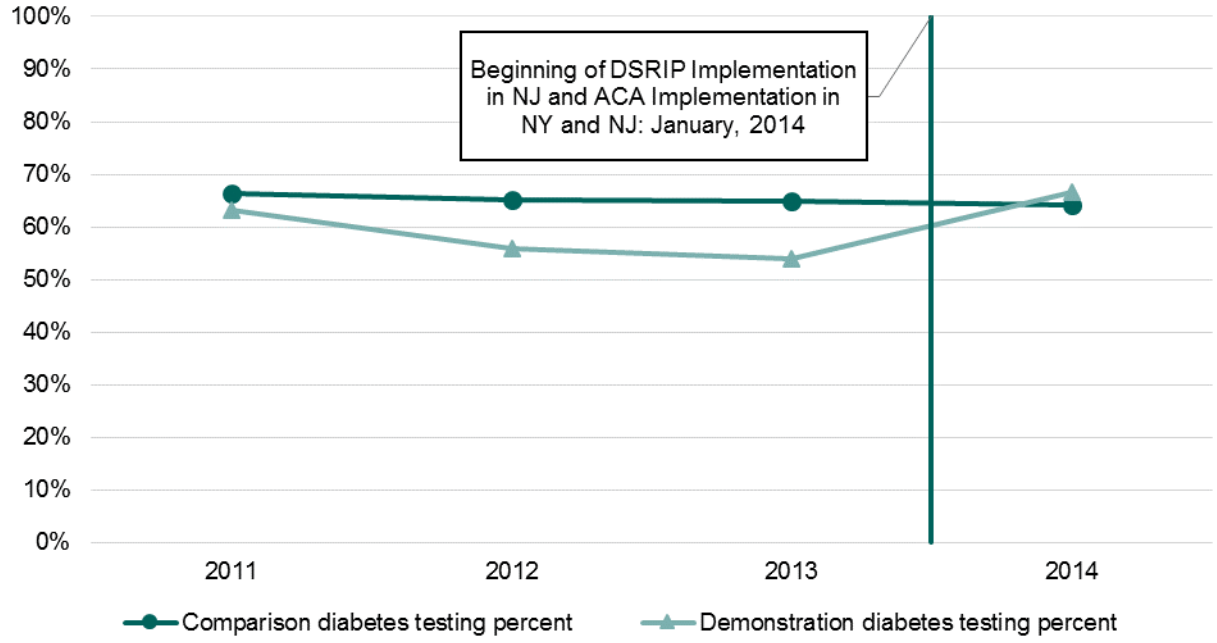
Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from New Jersey and New York, 2011 to 2014.

c. Diabetes testing

Figure III.7 shows the share of beneficiaries with diabetes who had an HbA1c test during the year in demonstration and comparison HSAs. In comparison HSAs, the share of beneficiaries with an HbA1c test remained relatively constant over the pre-period and into the post-period. By contrast, the share of beneficiaries living in demonstration HSAs with an HbA1c test decreased over the pre-period and increased in the year after implementation of DSRIP.

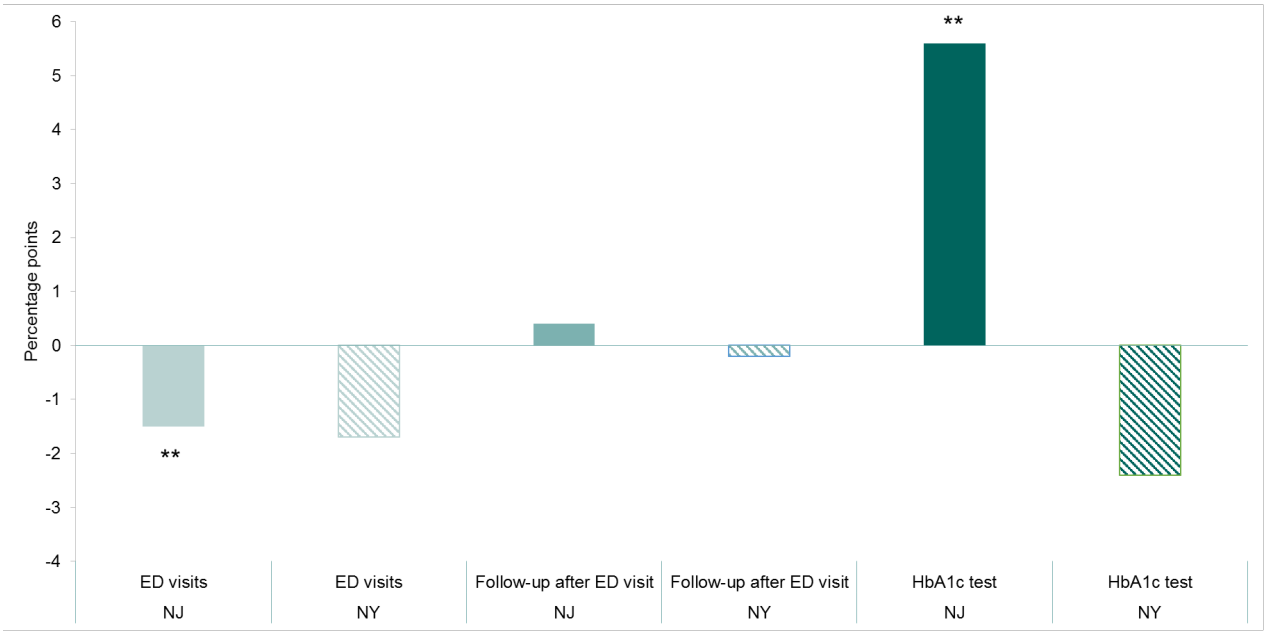
After controlling for individual and community-level characteristics, the probability of having a diabetes test increased in the demonstration group from the pre-period to the post-period. In the comparison group, the rate of diabetes test decreased during the same period. This suggests that the demonstration may have positively influenced diabetes testing (Figure III.8 and Table C.2). This result should be interpreted with caution for two reasons. First, there is a single data point in the post-period leading to less reliable estimates of the post-period rates. Second, the study population is younger in New Jersey than in New York, and rates of diabetes testing rise with age, meaning that the two groups may not be an ideal match.

Figure III.7. Diabetes testing in New Jersey demonstration and New York comparison groups by year, 2011–2014



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from New Jersey and New York, 2011 to 2014

Figure III.8. Impact of DSRIP in New Jersey between 2011 and 2014: Change from pre to post DSRIP



Source: Mathematica analysis of MAX and Alpha-MAX data from New Jersey and New York, 2011 to 2014.

Note: This figure presents findings based on a difference-in-differences analysis for three outcomes: emergency department [ED] visits, follow-up after ED visit for an ambulatory care sensitive coditions, and diabetes (HbA1c) testing. Each outcome was analyzed separately using a linear probability regression that adjusted for sex, age, clinical conditions, median household income, number of beds per resident, and number of hospitals per HSA. Each bar represents a difference between one of the adjusted outcomes before and after DSRIP for one of two groups: demonstration (or New Jersey DSRIP) HSAs and similar comparison (New York non-DSRIP) HSAs. ED visits and follow-up after an ED visit for ambulatory care sensitive conditions are measured at the person-quarter level. Diabetes testing is measured at the person-year level. * $p < 0.01$, ** $p < 0.001$.

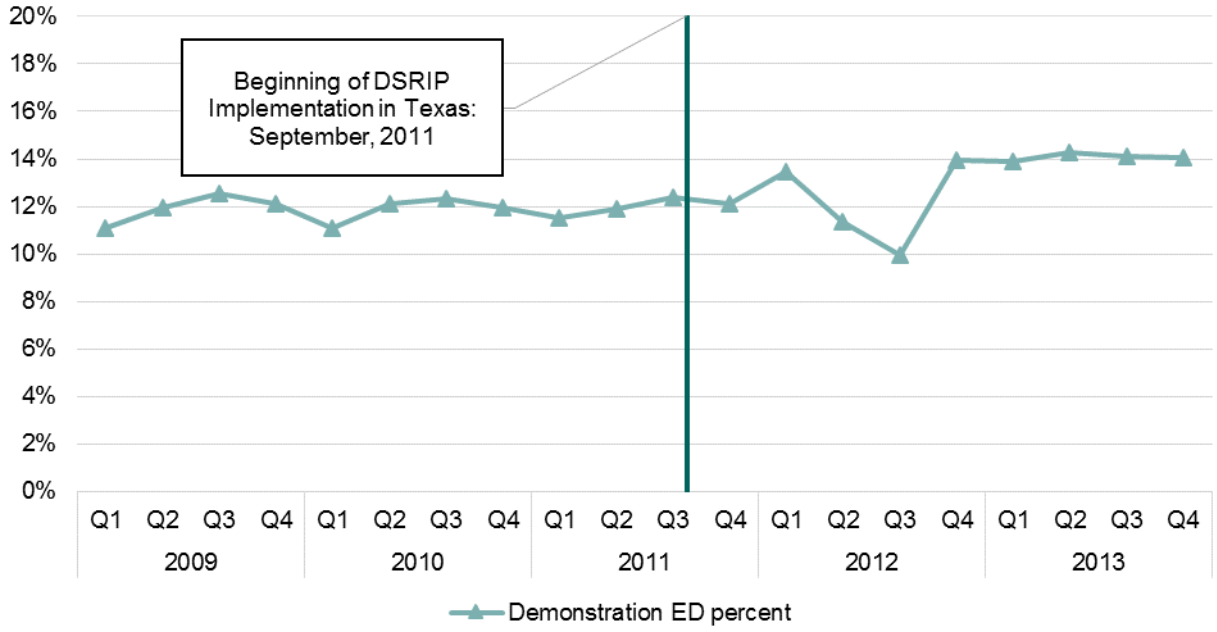
3. Texas

a. ED visits

Figure III.9 below shows the share of beneficiaries in Texas with an ED visit before and after implementation of DSRIP. In the pre-period, the share of beneficiaries with an ED visit appeared relatively stable (on average, roughly 12 percent) with some seasonal variation. This trend changed in 2012, with the prevalence of ED visits decreasing in the first two quarters and increasing in the fourth quarter. In 2013, the rate stabilized around 14 percent.

After controlling for individual and community-level characteristics, we found that the estimated trend in the probability of having an ED visit was flat, which is consistent with the unadjusted findings. Our estimates showed no change in the level of ED visits in the post-period; however, they did show a change in trend, with the trend in probability of having an ED visit increasing over time in the post-period (Figure III.12 and Table C.3).

Figure III.9. Share of beneficiaries with an emergency department visit in Texas by quarter, January 2009 through December 2013



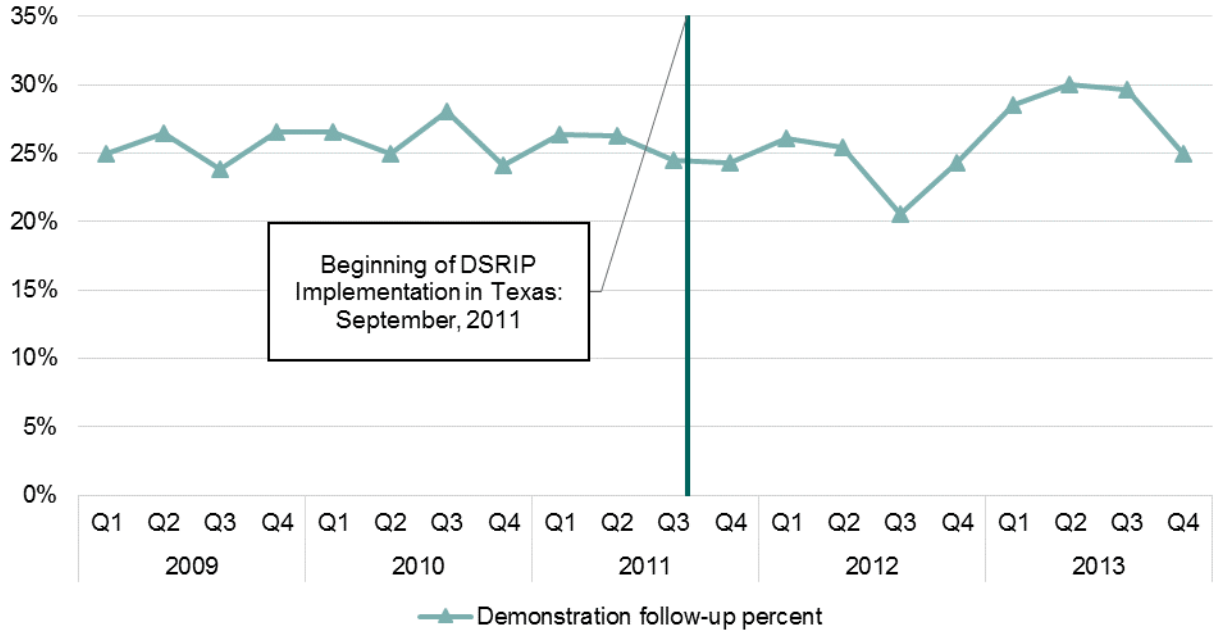
Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from Texas, 2009 to 2013.

b. Follow-up after an ED visit for an ACSC

The share of beneficiaries with a follow-up after an ED visit for an ACSC is shown in Figure III.10. There was a relatively stable trend (around 27 percent) in the pre-intervention period. In the post-period, we see considerably more variation in the trend over time.

After controlling for individual- and community-level characteristics, we found no significant results related to the probability of having a follow-up after an ED visit for an ACSC (Figure III.12 and Table C.3).

Figure III.10. Follow-up after an ED visit for an ACSC in Texas by quarter, January 2009 through December 2013



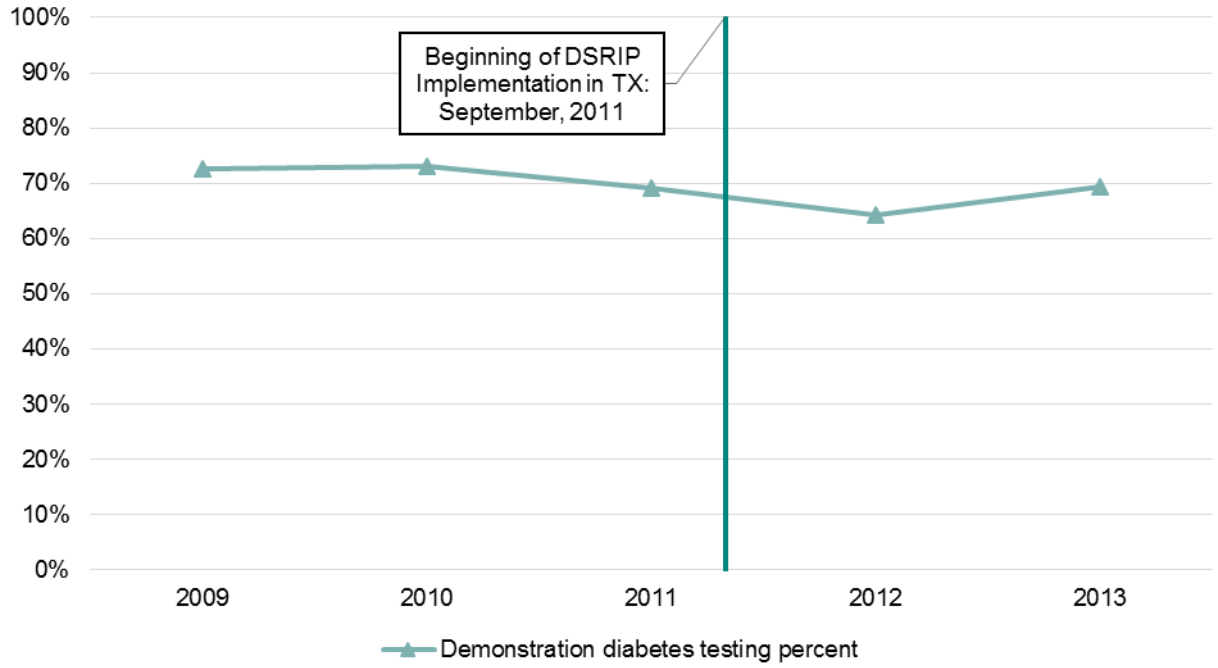
Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from Texas, 2009 to 2013.

c. Diabetes testing

Figure III.11 shows the share of beneficiaries with diabetes who had an HbA1c test during the year. The rate of testing decreased slightly in the pre-period and into the first year of the post-period, and increased slightly in the second year of the post-period.

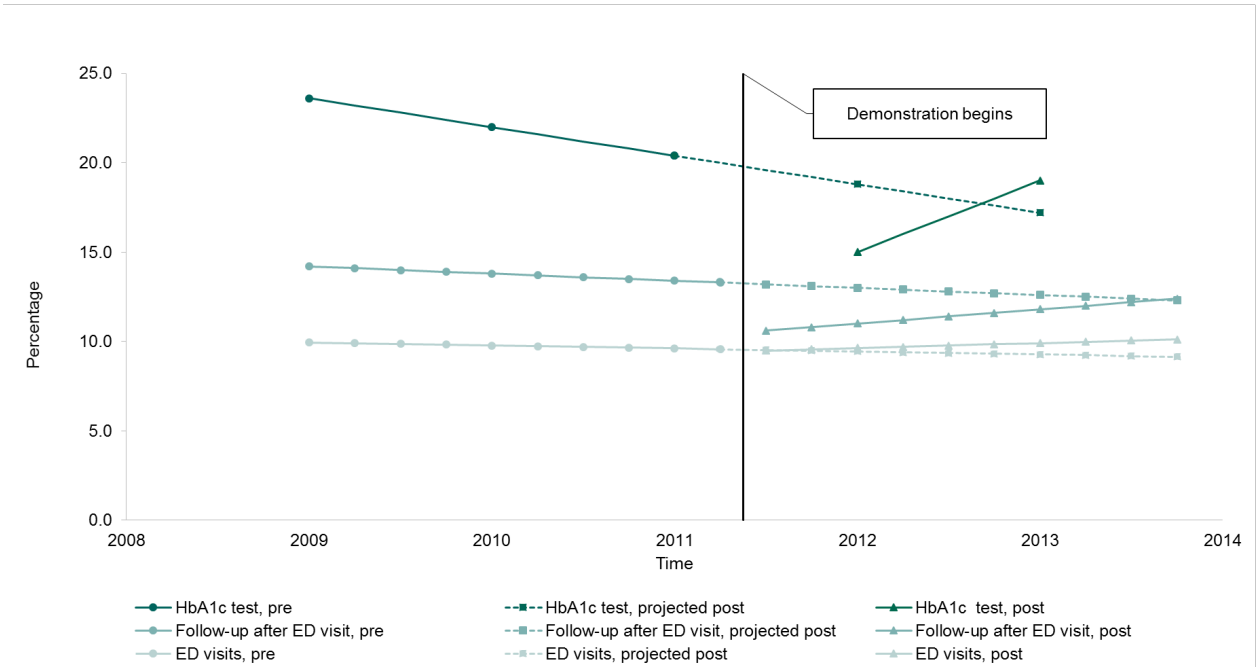
After controlling for individual and community-level characteristics, we found a downward trend in the probability of having a diabetes test in the pre-period. Immediately after DSRIP implementation, we found a decrease in the probability of having a diabetes test, followed by an increase in probability of having a diabetes test over time. (see Figure III.12 and Table C.3). However, these results on diabetes testing need to be interpreted with caution because of the limited data availability and lack of a comparison group in Texas.

Figure III.11. Diabetes testing in Texas by year, 2009–2013



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from Texas, 2009 to 2013.

Figure III.12. Impact of DSRIP in Texas between 2009 and 2013: Adjusted patient outcomes pre and post DSRIP



Source: Mathematica analysis of MAX and Alpha-MAX data from Texas, 2009 to 2013.

Note: This figure presents findings based on a simple interrupted time series analysis for three outcomes: emergency department [ED] visits, follow-up after ED visit for an ambulatory care sensitive coditions, and diabetes (HbA1c) testing. Each outcome was analyzed separately using a linear probability regression that adjusted for sex, age, clinical conditions, median household income, number of beds per resident, and number of hospitals per HSA. ED visits and follow-up after an ED visit for ambulatory care sensitive conditions are measured at the person-quarter level. Diabetes testing is measured at the person-year level.

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

IV. DISCUSSION

One purpose of the interim outcomes evaluation was to test the feasibility of the analytic strategy. The results showed that the demonstration and comparison groups usually had similar baseline trends in the outcomes of interest, which suggests that our strategy for identifying comparison groups in California and New Jersey was successful. Although the baseline trends tended to be parallel in New Jersey and New York, they did deviate at certain points. Further, the distribution of some covariates, such as age, varied across the two states. Given this variation, we will explore a multi-state comparison group in the final outcomes evaluation.

The second purpose of the interim outcomes evaluation was to generate preliminary findings. Because only limited data were available in the period after DSRIP was implemented, these preliminary findings should be interpreted with caution. At this early stage, the analyses yielded mixed results across measures and states:

- In California, the probability of having an ED visit decreased in both the demonstration and comparison groups in the post-period (a positive result), but there was a smaller decrease in the demonstration group (that is, DSRIP was associated with a relatively worse outcome). DSRIP was not associated with having a follow-up after an ED visit for ACSC. The probability of having a diabetes test decreased for both groups (a negative result), though there was a smaller decrease in the demonstration group (that is, DSRIP was associated with a relatively better outcome).
- In New Jersey, the DSRIP program demonstrated mixed effects. The probability of having an ED visit decreased across both demonstration and comparison groups in the post-period (a positive result), but there was a smaller decrease in the demonstration group (that is, DSRIP was associated with a relatively worse outcome). DSRIP was not associated with having a follow-up after an ED visit for ACSC. The probability of having a diabetes test after DSRIP implementation increased for the demonstration group and decreased for the comparison group, a positive result for the DSRIP demonstration.
- In Texas, the DSRIP program demonstrated mixed effects. The probability of having an ED visit increased over time after DSRIP implementation (a negative result). The probability of having a follow-up after an ED visit for ACSC was not associated with DSRIP. We found a decrease in the probability of having a diabetes test immediately after DSRIP implementation, followed by an increase in the probability of having a diabetes test over time during the post-period (a mixed result).

A. What have other studies found?

Overall, these preliminary results are consistent with our rapid cycle report findings, which suggested that the goals of DSRIP in the early years of implementation, particularly in California and Texas, were focused more heavily on infrastructure development as opposed to service redesign and delivery system transformation. The focus on infrastructure development was seen as a critical first step toward improving clinical quality and eventually clinical and population outcomes in later years of the program. With this context, it might be expected that early demonstration effects on clinical quality, particularly on measures related to ED use, are minimal.

These findings also add to a small body of research that has examined the impact of DSRIP programs and that, to date, has been limited to state-based evaluations. These include the final program evaluations in California (Pourat et. al. 2016), and Texas (Texas Health and Human Services 2017), and the midpoint evaluation in New Jersey (Chakravarty et. al. 2015). Each state-based evaluation, however, has features that preclude us from directly comparing our results to the findings of the study.

The final evaluation in California examined the impact of DSRIP over the entire performance period from 2011 to 2015. To assess the impact of DSRIP on quality of care, patient outcomes, and cost, the study relied on DSRIP semiannual and annual reports that were based on data reported by the DPHs, as well as provider-level survey data. Overall, the assessment of semiannual and annual reports showed that the DPHs achieved nearly all milestones (97 percent) associated with the program, and exceeded the annual targets for improvement milestones. However, DPHs tended to set targets they could easily achieve in order to ensure receipt of DSRIP funds. As a result, these findings do not necessarily suggest that DPHs achieved significant improvements in delivery system transformation or quality of care relative to what would otherwise have occurred. The survey data provide information about DPHs' perceptions of the project's impacts. DPHs perceived that DSRIP projects have a high impact on quality of care, patient outcomes, and cost, but additional data were not provided to support these results. Given the design of this evaluation, it is complementary to our project but cannot be used to directly confirm our findings.

The midpoint assessment in New Jersey, which examined the impact of DSRIP from 2011 to 2013,[15] relied on an analysis of Medicaid claims data to examine patient care, health, costs, and hospital finances. Although the midpoint assessment used an approach similar to this study in its data sources and analytic strategy, the analysis only included data from 2011 through 2013, which was before DSRIP implementation. As such, the results are useful for understanding changes in outcomes during the DSRIP planning period, but should not be interpreted as estimates of the impact of the DSRIP program, which our interim and final evaluations aim to produce. Furthermore, the midpoint assessment relied on a different set of measures to assess program impact. In spite of this, the results do provide helpful context about trends in patient care and health outcomes before DSRIP implementation. When comparing data from 2013 to 2011 and 2012, improvements took place in some measures (for instance, avoidable hospitalizations for asthma and diabetes), others worsened (for instance, ED visits for asthma among adults), and many had no changes (for instance, follow-up after a hospitalization for mental illness and 30-day readmissions for heart failure, acute myocardial infarction, pneumonia, and chronic obstructive pulmonary disease).

The final evaluation in Texas, which examined the impact of DSRIP from 2013 to 2016, focused on the effects of one specific DSRIP project—care navigation—on the quality of care, patient health, and costs. The state did not find that patients receiving care from sites implementing care navigation projects had better outcomes than patients at other sites. In addition to the final evaluation report, the state submitted a companion document that examined the success rate for DSRIP outcomes over the same time period (Smith 2017). The success rate

[15] Although the New Jersey DSRIP program planning period began in 2013, implementation did not begin until 2014.

was defined as the percentage of pay-for-performance outcomes on which providers earned at least partial payment.[16] Hospitals participating in DSRIP demonstrated a 79 percent success rate for the first performance year. Across all DSRIP providers, the success rate for ED visits per 100,000 individuals (Medicaid beneficiaries and uninsured) was 63 percent, ED visits for ACSCs was 66 percent, and ED visits for diabetes was 93 percent. Although not directly comparable to our results, the patterns of success rates correspond with our current findings. For instance, ED visits per 100,000 individuals had the lowest success rate, and ED visits for diabetes had the highest success rate.

B. Study limitations

Our evaluation of the effect of DSRIP on quality and efficiency of care faced several methodological challenges including (1) a short follow-up period, due to data limitations, (2) concurrent changes in Medicaid policy and the composition of the Medicaid population, (3) innate differences between the providers that participated in DSRIP and other providers, (4) a lack of information about outcomes for uninsured individuals in the study data, and (5) the need to use simple evaluation models, due to the short follow-up period and the lack of credible comparison groups. As described below, we addressed these challenges to the extent possible in the interim evaluation and will further strengthen the design for the final evaluation. Nevertheless, given these challenges, this report should be primarily viewed as a proof of design and an indicator of baseline trends with results that are preliminary. Fully assessing the demonstration's impact will require a longer data period that will be available for the final evaluation.

One of our most significant limitations was the short follow-up period available for this interim evaluation, which limits our ability to observe longer-term effects, as well as the types of models we can estimate. Medicaid administrative data provide the most complete record of services rendered to Medicaid beneficiaries, but are not currently available beyond 2013 or 2014, which limits the post-period we can observe. This is particularly problematic for New Jersey, in which there is only one year of data after implementation of DSRIP, and for the diabetes measure, which is measured annually rather than quarterly. We anticipated that the demonstration would not affect outcomes immediately; therefore, the results of this report should be considered preliminary, and null results in several cases are unsurprising. In addition, the short follow-up periods limit the types of models that can be used and make it difficult to estimate the parameters of interest accurately. Because the study period in the final evaluation will be longer, we are optimistic that it will be sufficient for identifying the demonstration's impact across all the domains of clinical quality, readiness for value-based purchasing, population health, and total cost of care. Further, the extended study period should enable us to estimate demonstration effects using more rigorous methods, such as a comparative interrupted time series approach.

In addition, the demonstration is unfolding in the context of a rapidly changing health system, and many forces beyond the demonstration affect the outcomes of interest. This makes isolating the impact of DSRIP a challenge. To account for this, to the extent possible, we selected comparison groups that were affected by similar forces and used models that

[16] To receive partial payment, providers must achieve 25 percent of the performance goal for a given year.

incorporated a robust set of covariates to capture measurable changes in the environment. We plan to improve upon these methods for the final evaluation. For instance, we plan to add a comparison group for Texas, improve our comparison group for New Jersey if possible, and rely on models that allow for differing trends between the demonstration and comparison groups. Despite these strategies, teasing out the effects of DSRIP in a time of health system change remain a difficult endeavor.

A third limitation is that DSRIP HSAs differ from the non-DSRIP HSAs on some factors that we were unable to control for in our analysis. In California, the DSRIP program was implemented only in designated public hospital systems, while non-DSRIP HSAs are served by other types of hospitals. In the evaluation of the New Jersey DSRIP program, the DSRIP and non-DSRIP HSAs are located in different states. As a result, DSRIP and non-DSRIP hospitals and the communities that they serve may differ on a number of unobserved factors, including regulations, policies, and practices, which may also differentially affect the outcomes observed these. However, by matching demonstration hospitals to comparison hospitals on observable attributes and reviewing baseline trends across both groups, we sought to maximize the validity of the comparison.

Fourth, because Medicaid administrative data were used, the interim evaluation lacks information about care delivered to the uninsured individuals, a critical target population of the DSRIP programs in California, New Jersey, and Texas. The final evaluation will consider additional data sources, such as discharge data from the Healthcare Cost Utilization Project, which will include outcome data for on the uninsured population.

Finally, the ideal approach to an evaluation is a randomized trial, and any observational study requires simplifying assumptions. For example, the difference-in-differences model, used for California and New Jersey, is valid if both the demonstration and comparison groups have identical underlying trends. If this assumptions is not satisfied, then the estimated impact may be confounded by the differential trend between groups. Similarly, the interrupted time series model used in Texas is valid if the demonstration group has a consistent, underlying trend in both the pre and post periods, but, again, if the assumption is not satisfied, then the estimated impact may not be valid. In the final evaluation, we plan to use a comparative interrupted time series model for all states, a model that accommodates trends that differ between the demonstration and comparison group and between the pre and post period. Results based on the current models and associated time periods must be viewed as preliminary, and even results from future comparative interrupted time series will only be valid if key assumptions are met.

C. Future research

In the final outcomes report, which is scheduled for 2019, we plan to include additional states that are implementing DSRIP or DSRIP-like demonstrations, such as Massachusetts, New Hampshire, New York, and Washington.[17] We will also explore identifying a multi-state comparison group for Texas—and potentially, New Jersey—to improve the rigor of the evaluation. We expect the final analyses to include additional years of data, which will allow us to estimate comparative interrupted time series models. We also plan to include additional

[17] Due to limitations with the MAX data, we will have to obtain hospital discharge data from the Healthcare Cost and Utilization Project to include some states such as Massachusetts in the final outcomes evaluation.

analyses of the effect of DSRIP on delivery system transformation (for instance, follow-up after an ED visit for mental illness); clinical quality (for instance, hospital readmissions after 30 days); preparation for value-based payment (for instance, participation in Medicare alternative payment methods); population health (for instance, controlled diabetes); and total cost of care (for instance, total cost per beneficiary per month). Finally, we anticipate incorporating inpatient and emergency department discharge data from the Hospital Cost and Utilization Project, enabling us to assess outcomes related to hospital use for uninsured individuals.

The outcomes evaluation, coupled with the national evaluation team's rapid-cycle reports focused on demonstration implementation, will help both CMS and state officials understand the influence of DSRIP programs on delivery system transformation, clinical quality, preparation for value-based purchasing, population health, and cost of care for Medicaid beneficiaries and the uninsured. The ultimate goal of the evaluation is to give policymakers better tools for reforming the delivery system to support the Triple Aim principles of better care, improved health, and lower costs.

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

REFERENCES

Baller, Julia, Marian Wrobel, Natalya Verbitsky Savitz, Mariel Finucane, and Carol Irvin. "Delivery System Reform Incentive Payments: Design Supplement." 2017. Available at https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/eval-dsgn-dlvry-incntv-pymnts.pdf.

Centers for Medicare & Medicaid Services, and California Health and Human Services Agency. "Special Terms and Conditions, No. 11-W-00193/9, California Bridge to Reform Demonstration." As amended February 27, 2015.

Centers for Medicare & Medicaid Services, and 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 2017 Reporting." Baltimore, MD: Centers for Medicare & Medicaid Services, June 2017. Available at https://www.medicaid.gov/medicaid/quality-of-care/downloads/medicaid-and-chip-child-core-set-manual.pdf.

Centers for Medicare & Medicaid Services, and 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 2017 Reporting." Baltimore, MD: Centers for Medicare & Medicaid Services, June 2017. Available at https://www.medicaid.gov/medicaid/quality-of-care/downloads/medicaid-adult-core-set-manual.pdf.

Centers for Medicare & Medicaid Services, and Texas Health and Human Services Commission. "Special Terms and Conditions, No. 11-W-00278/6, Texas Healthcare Transformation and Quality Improvement Program." As amended February 26, 2015.

Chakravarty, Sujoy, Joel C. Cantor, James T. Walkup, and Jian Tong. "A Midpoint Evaluation of the New Jersey DSRIP Program: Findings from Stakeholder Interviews, Hospital Survey, Medicaid Claims Data, and Reported Quality metrics." 2014. Available at http://www.state.nj.us/humanservices/dmahs/home/NJCW_Renewal_App_C2_DSRIP_Mid_Point_Evaluation_Report.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.

Pourat, Nadereh, Ying-Ying Meng, Arleen Leibowitz, Jack Needleman, Xiao Chen, Dylan H. Roby, Max Hadler et al. "Final Evaluation Report of California's Delivery System Reform Incentive Payments (DSRIP) Program." 2016.

Smith, Charles. "Texas DSRIP Outcomes: Companion to the Final Evaluation Report of the 1115(a) Texas Demonstration Waiver—Healthcare Transformation and Quality Improvement." 2017. Available at https://hhs.texas.gov/sites/default/files//documents/laws-regulations/policies-rules/1115-docs/Evaluation-Companion-Document.pdf.

State of New Jersey Department of Health. "Delivery System Reform Incentive Payment Program Planning Protocol." Trenton, NJ: State of New Jersey Department of Health, August 2013. Available at https://dsrip.nj.gov/Documents/NJ%20DSRIP%20PLANNING%20PROTOCOL_v1_08-09-2013.pdf.

Texas Health and Human Services. "Evaluation of the 1115(a) Texas Demonstration Waiver—Healthcare Transformation and Quality Improvement." May, 2017. Available at https://hhs.texas.gov/sites/default/files/documents/laws-regulations/policies-rules/1115-docs/tool-guidelines/Evaluation-Texas-Demonstration-Waiver.pdf.

APPENDIX A

DSRIP PROGRAM CHARACTERISTICS

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

Table A.1. DSRIP program characteristics in three states

Characteristic	California[a]	Texas[b]	New Jersey[b]
Approval date	11/1/2010	12/12/2011	10/1/2012
Expiration date	12/31/2020	9/30/2021	6/20/2020
Total program funding	$14.135B	$23.793B	$1.083B
Additional funding introduced by DSRIP?	Yes	Yes	No
Type of providers eligible to receive incentive payments	Designated public hospital systems and district/ municipal public hospitals	Regional consortia of providers	Acute care hospitals
Number of providers	55 hospitals	338 providers in 20 Regional Healthcare Partnerships	49 hospitals
Number of projects	221[c]	1,450	49

Sources: U.S. Centers for Medicare & Medicaid Services and California Health and Human Services Agency. "Special Terms and Conditions, No. 11-W-00193/9, California Bridge to Reform Demonstration." Approval Period: November 1, 2010 through October 31, 2015, as amended February 27, 2015.

U.S. Centers for Medicare and Medicaid Services and New Jersey Department of Human Services Division of Medical Assistance and Health Services. "Special Terms and Conditions, No. 11-W-00279/2 (Title XIX), New Jersey Comprehensive Waiver Demonstration." Approval Period: October 1, 2012 through June 30, 2017, technical corrections approved August 14, 2014.

U.S. Centers for Medicare and Medicaid Services and New Jersey Department of Human Services Division of Medical Assistance and Health Services. "Special Terms and Conditions, No. 11-W-00279/2 (Title XIX), New Jersey Comprehensive Waiver Demonstration." Approval Period: August 1, 2017 through June 30, 2022, amended October 31, 2017.

U.S. Centers for Medicare and Medicaid Services and Texas Health and Human Services Commission. "Special Terms and Conditions, No. 11-W-00278/6, Texas Healthcare Transformation and Quality Improvement Program." Approval Period: December 12, 2011 through September 30, 2016; amendment approved February 26, 2015.

U.S. Centers for Medicare and Medicaid Services and Texas Health and Human Services Commission. "Special Terms and Conditions, No. 11-W-00278/6, Texas Healthcare Transformation and Quality Improvement Program." Approval Period: January 2, 2018 through September 30, 2022; approved December 21, 2017.

[a] Program currently in renewal period.

[b] Program in extension period.

[c] Number of projects in first waiver period.

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

Area of clinical focus[a]	California % of DPHs carrying out project	Texas % of RHPs carrying out project	New Jersey % hospitals carrying out project
Primary care	100	100	-
Appropriate care in appropriate settings	29	100	22
Diabetes care	-	-	24

Sources: Mathematica analysis of California Health Care Safety Net Institute. "Aggregate Public Hospital System Annual Report on California's 1115 Medicaid Waiver's Delivery System Reform Incentive Payment Program: Demonstration Year 8." (Version 2). Oakland, CA: California Health Care Safety Net Institute, December 31, 2013.

State of New Jersey, Department of Health. "DY3 New Jersey DSRIP Hospitals." Available at https://dsrip.nj.gov/Documents/Demonstration%20Year%20Three%20Hospital%27s%20Participation%20List.xlsx.

Texas Health and Human Services Commission. "DSRIP Approved and Active Projects as of January 6, 2016 [Excel file]." Available at http://www.hhsc.state.tx.us/1115-docs/011316/Active-DSRIP-Projects-with-Cat-3-20160106.xls.

[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.

- indicates project type was not included in state project menu

DPH = designated public hospital system; RHP = Regional Healthcare Partnership

APPENDIX B

METHODS

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

Appendix B provides additional detail about the methods used in the draft interim outcomes evaluation. Appendix Table B.1 provides information about the three outcome measures, including the NQF number, whether it is in a Medicaid Core Set, the clinical focus area represented, and specifications for calculating the numerator and denominator. Appendix Figure B.1 portrays our approach to selecting the eligible population for outcome measures. Appendix Figure B.2 depicts the framework used to determine the appropriate comparison group for each study state. Appendix Figures B.3 to B.5 illustrates the sample size in the demonstration and comparison groups in California, New Jersey and New York, and Texas, respectively. Appendix Section B.1 describes our approach for hospital service area matching, with results presented in Appendix Tables B.2 and B.3. Appendix Table B.4 describes MAX data availability for each study state. Finally, Appendix Section B.2 describes how we estimated the demonstration effects.

Table B.1. Outcome measures used in the interim evaluation

Outcome measure	NQF number	In a Medicaid core set	Clinical focus area(s)	Numerator	Denominator
ED visits	n.a.	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	n.a.	No	Primary care	Eligible adults with an outpatient visit within 7 days of discharge from the ED for asthma, COPD, hypertension or diabetes	Adults ages 18 to 64 with an ED visit for asthma, COPD, hypertension, or diabetes[b]
Comprehensive diabetes care: HbA1c control	NQF 0059	Yes	Diabetes care	Eligible adults with HbA1c testing	Adults ages 18 to 64 with diabetes [c]

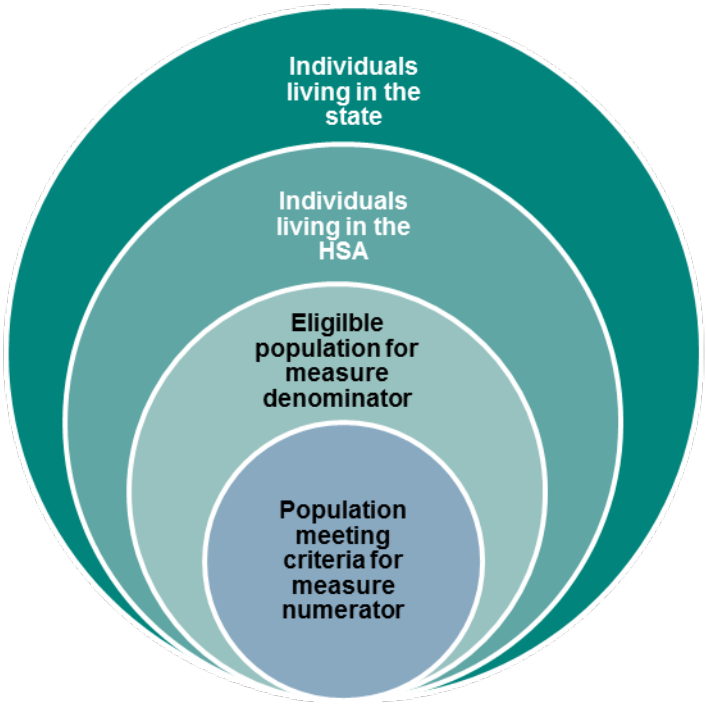
[a] The Medicaid core set of child quality measures includes the measure Ambulatory Care—ED visits for beneficiaries, ages from birth to age 21. We have adapted this measure for the adult beneficiaries ages 18 to 64. Most notably, the measure excludes ED visits for mental illness, and alcohol and other 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 before.

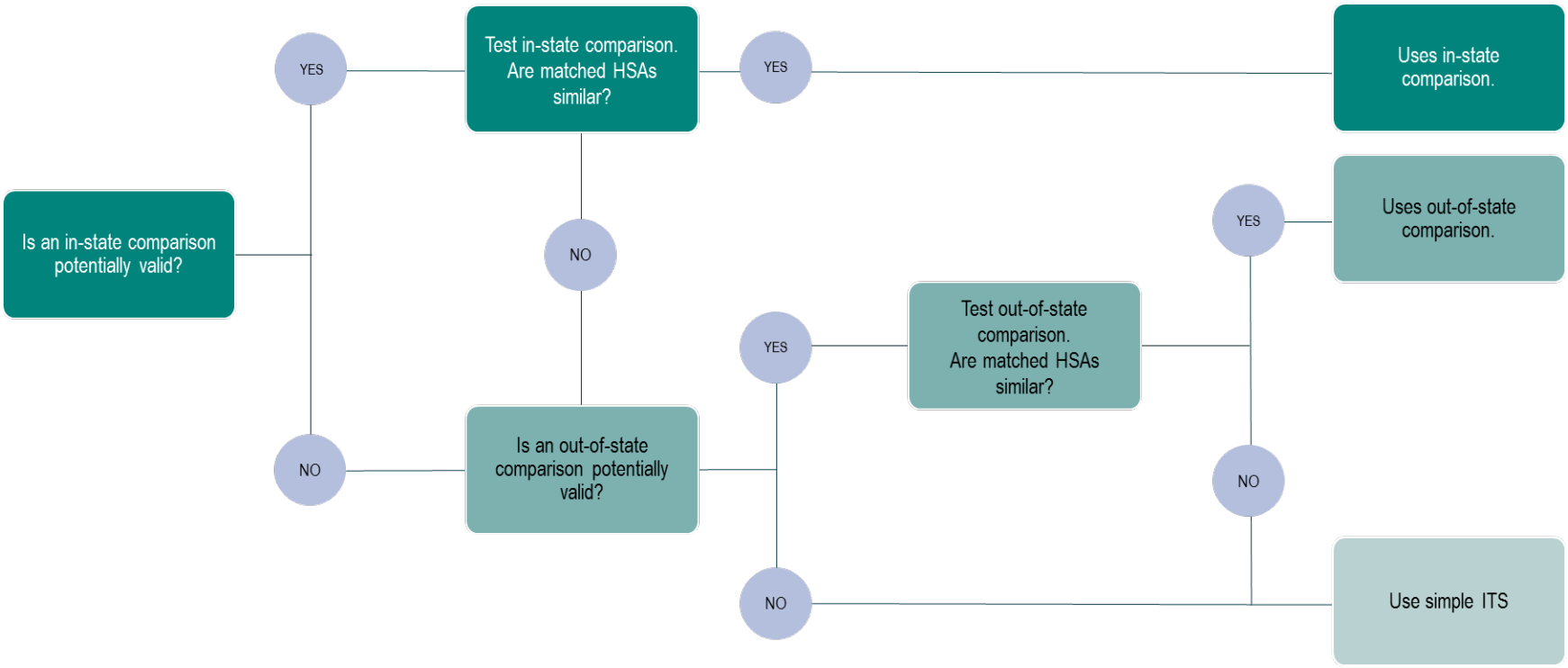
COPD = chronic obstructive pulmonary disease; ED = emergency department; n.a. = not applicable; NQF = National Quality Forum.

Figure B.1. Eligible population for outcome measures



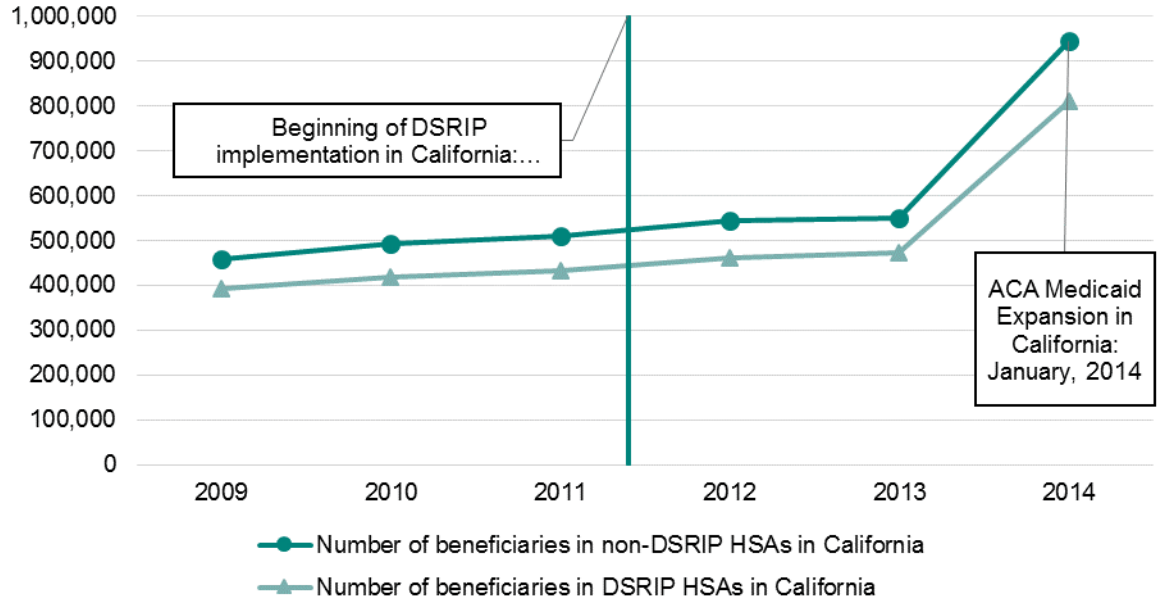
HSA = hospital service area.

Figure B.2. Selection of the comparison group



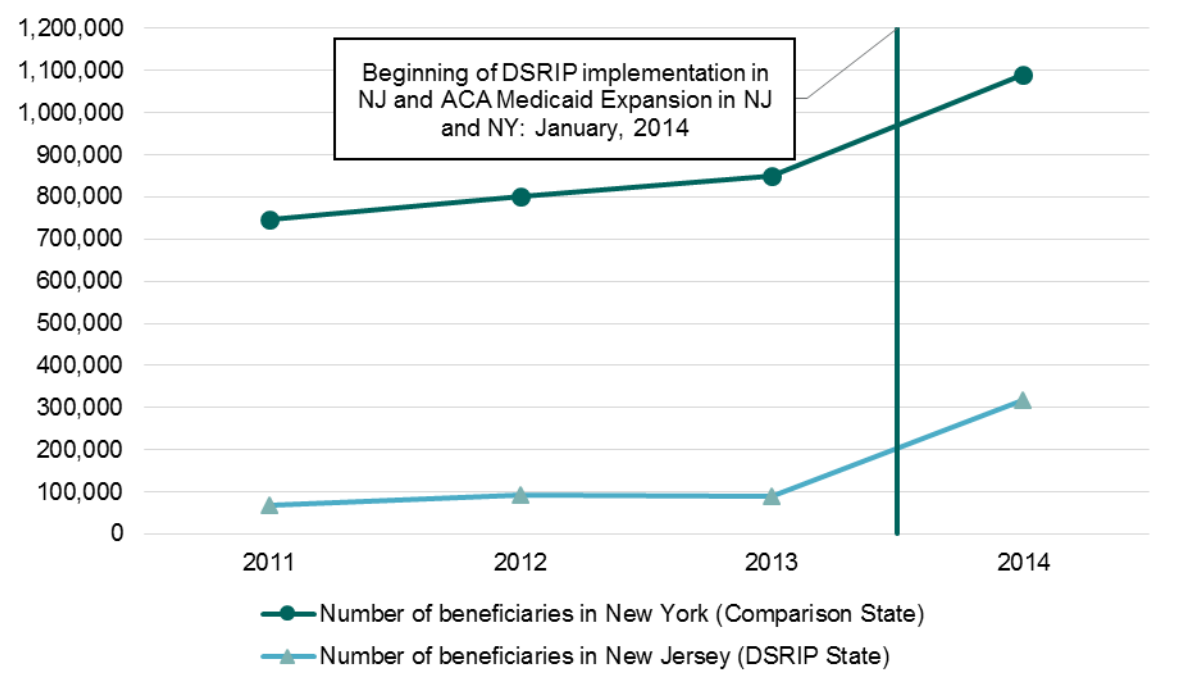
HSA = hospital service area; ITS = interrupted time series.

Figure B.3. Medicaid study sample size in demonstration and comparison groups in California by quarter, January 2009 to December 2014



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from California, 2009 to 2014.

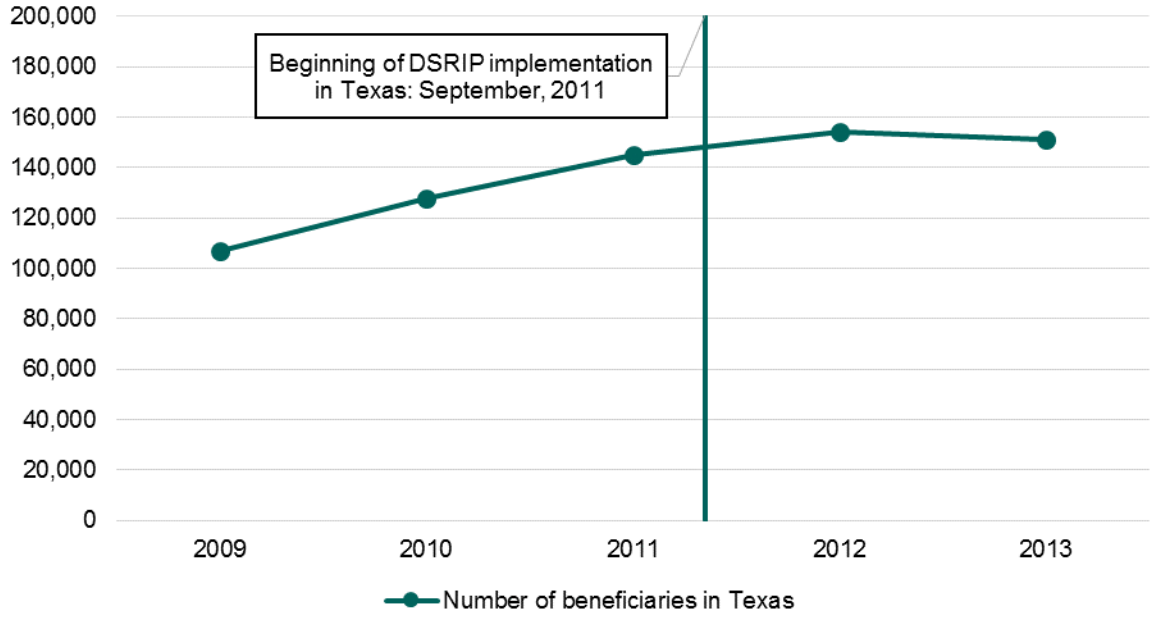
Figure B.4. Medicaid study sample size in New Jersey demonstration and New York comparison groups by quarter, January 2009 to December 2014



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from New Jersey and New York, 2011 to 2014.

Note: DSRIP implementation and ACA expansion are on the same date (Jan 2014), but are not aligned on the graph. Listed DSRIP implementation date is for NY, not NJ.

Figure B.5. Medicaid study sample size in Texas by quarter, January 2009 to December 2013



Source: Mathematica analysis of unadjusted MAX and Alpha-MAX data from Texas, 2009 to 2013

SECTION B.1. HOSPITAL SERVICE AREA MATCHING

In order to create a comparison group that was similar to the demonstration group, we matched each demonstration HSA to one or more eligible comparison HSAs. We employed an exact matching strategy in which we choose four variables considered to be predictors of whether or not a hospital participated in DSRIP. Those four variables are listed here in order of importance.

- **Percentage of all discharges in the HSA that were Medicaid discharges**. The source for this data was CMS' Healthcare Cost Report Information System's IME_GME files.[18] In order to be considered a match on this variable, a demonstration and comparison HSA had to be in the same "category" for Medicaid discharges—a high, medium, or low number of Medicaid discharges. An HSA is considered to have high numbers of Medicaid discharges if the percentage of spending by Medicaid on their total discharges at least one standard deviation higher than the mean for all HSAs in the pool of eligible comparison and demonstration HSAs. Conversely, the number of Medicaid discharges is considered low if the percentage of spending by Medicaid on the HSA's total discharges is at least one standard deviation lower than the mean for all HSAs in the pool of eligible comparison and demonstration HSAs. All HSAs outside these two categories are considered to have a medium level of Medicaid discharges.
- **Average size of hospitals in each HSA (categorical number of beds)**: The source for this variable was the 2009 American Hospital Association survey. We recoded the American Hospital Association's Bed Size Code (a categorical variable for number of beds) based on the average number of beds in each hospital within each HSA, weighted by each hospital's number of inpatient admissions. In order to be considered a match on this variable, demonstration and comparison HSAs had to have the same value for the recoded Bed Size Code.
- **Core Based Statistical Areas (CBSA) Type of the HSA (rural, urban, or mix)**: The source for this variable was the 2009 American Hospital Association survey. In order to be considered a match, the HSAs had to have the same CBSA type. If an HSA had hospitals with two different CBSA types (which happened infrequently), then the HSA was called a mix and could only be matched to another mixed HSA.
- **Category of Medicare ACSC discharges per 1,000 Medicare beneficiaries**: The source for this variable was Dartmouth Atlas's Selected Measures of Primary Care Access and Quality[19] from 2009. In order to be considered a match on this variable, two HSAs had to be in the same category of ACSC discharges, which we defined in the same manner as we did the Medicaid discharges, described above for the first variable.

[18] Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Cost-Reports/Hospital-2010-form.html. Accessed July 31, 2017.

[19] Available at http://www.dartmouthatlas.org/tools/downloads.aspx?tab=35#primary. Accessed July 31, 2017.

We matched demonstration HSAs in California with non-demonstration HSAs in California, and matched demonstration HSAs in New Jersey with HSAs in New York. A demonstration HSA is one in which at least one hospital participated in DSRIP.

Of the 17 demonstration HSAs in California, 16 could be matched on all four matching variables, and one was matched on the first three of the four matching variables. Of the 41 demonstration HSAs in New Jersey, 37 matched on all four variables, and the remaining four matched on the first three of the four matching variables. Therefore, all eligible demonstration HSAs were successfully matched to at least one comparison HSA. See Table B.2 for full matching results, including those for each demonstration HSA and all of the comparison HSAs that were matched to it. The median number of comparison HSAs matched to each demonstration HSA in New Jersey was 3. Several demonstration HSAs in New Jersey had only one matched comparison HSA. Trenton had the maximum number of matched comparisons in this set, with 22 matched comparison HSAs—a clear outlier. The median number of comparison HSAs matched to each demonstration HSA in California was 4. As in New Jersey, several demonstration HSAs in California had only one matched comparison HSA. San Jose had the maximum number of matched comparisons in this set, with 12 matched comparison HSAs.

We tested for balance by examining standardized differences (SD) before and after the matching process. The SD statistic compares means of covariates between demonstration and comparison HSAs, standardized by the pooled standard deviation of the demonstration and comparison HSAs. For the standardized difference before matching, the difference is computed between all HSAs in a given demonstration state and all HSAs available to be matched to the HSAs in that state. The SD after matching is the difference between the mean of matched demonstration HSAs and the mean of matched comparison HSAs.[20] An SD of less than 0.25 is generally considered to be a good match.[21] Although not all characteristics and outcomes we used to test for balance met this standard after matching, our standardized differences generally improved after matching. Therefore, we believe this matched sample is an improvement over the unmatched sample for this interim evaluation and we expect the match quality to improve substantially for the final evaluation.

[20] Methodology was modeled on balance testing methodology on page 115 of the Partnership For Patients evaluation report: https://downloads.cms.gov/files/cmmi/pfp-interimevalrptapp.pdf. Accessed August 23, 2017.

[21] What Works Clearinghouse. Procedures and Standards Handbook. Version 3.0. Available at: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf. Accessed July 23, 2017.

Table B.2. Results of the balance testing

HSA characteristic or outcome variable	Mean of demonstration HSAs before matching	Mean of comparison HSAs before matching	Standardized difference before matching	Mean of matched demonstration HSAs after matching	Mean of matched comparison HSAs after matching	Standardized difference after matching
California						
Percentage of all discharges in the HSA that were Medicaid discharges	0.22	0.16	0.53	0.22	0.18	0.38
Average size of hospitals in each HSA (number of beds)	327.66	174.82	1.17	327.66	310.59	0.13
Core Based Statistical Areas (CBSA) Type of the HSA (rural, urban, or mix)	71.17	73.93	-0.37	71.17	72.51	-0.18
Category of Medicare ACSC discharges per 1,000 Medicare beneficiaries, baseline rate	54.49	54.78	-0.02	54.49	55.84	-0.08
ED visits, baseline rate (in 2011 year)	-0.06	-0.04	-0.10	-0.06	-0.10	0.28
Follow-up after discharge from the ED for ACSCs (asthma, COPD, hypertension, and diabetes), baseline rate	-0.10	0.00	-0.11	-0.10	-0.51	0.46
Comprehensive diabetes care: HbA1c testing, baseline rate	0.05	-0.03	0.54	0.05	0.03	0.15
New Jersey						
Percentage of all discharges in the HSA that were Medicaid discharges	0.06	0.10	-0.64	0.06	0.09	-0.40
Average size of hospitals in each HSA (number of beds)	340.04	254.52	0.43	340.04	319.51	0.10
Core Based Statistical Areas (CBSA) Type of the HSA (rural, urban, or mix)	72.03	76.50	-0.62	72.03	74.09	-0.29
Category of Medicare ACSC discharges per 1,000 Medicare beneficiaries	73.20	77.06	-0.21	73.20	68.80	0.24
ED visits, baseline rate (in 2013)	0.07	0.02	0.15	0.07	0.00	0.22
Follow-up after discharge from the ED for ACSCs (asthma, COPD, hypertension, and diabetes), baseline rate	-0.07	0.05	-0.15	-0.07	-0.08	0.02
Comprehensive diabetes care: HbA1c testing, baseline rate	-0.11	-0.03	-0.35	-0.11	-0.08	-0.14

Notes: We matched on (1) percentage of all discharges in the HSA that were Medicaid discharges; (2) average size of hospitals in each HSA (number of beds); (3) Core Based Statistical Areas (CBSA) type of the HSA (rural, urban, or mix); and (4) category of Medicare ACSC discharges per 1,000 Medicare beneficiaries. Emergency department (ED) visit rate, follow-up after discharge from ED for ACSCs, and diabetes testing were the three outcomes of interest. We did not use the outcome measures in our matching strategy. In the final interim outcomes evaluation report, we plan to include baseline rates in the three outcome measures as a part of our matching strategy.

The full list of demonstration and matched comparison HSAs is shown in Table B.3 below. The comparison HSAs matched to a given demonstration HSA have the same "Demonstration/Comparison Set Index" number as that demonstration HSA and are listed as "Matched." The demonstration HSAs are listed as "Demonstration."

Table B.3. Demonstration and matched comparison HSAs

New Jersey/New York			California		
Demonstration/ comparison set index	Demonstration or matched comparison HSA	HSA name	Demonstration/ comparison set Index	Demonstration or matched comparison HSA	HSA name
1	Demonstration	Atlantic City, NJ	42	Demonstration	Alameda, CA
1	Matched	Buffalo, NY	42	Matched	Hayward, CA
1	Matched	Rochester, NY	42	Matched	Lakewood, CA
1	Matched	Albany, NY	42	Matched	Antioch, CA
2	Demonstration	Bayonne, NJ	42	Matched	San Dimas, CA
2	Matched	Bronxville, NY	42	Matched	Glendora, CA
2	Matched	Bethpage, NY	42	Matched	Bellflower, CA
2	Matched	Dobbs Ferry, NY	42	Matched	Los Alamitos, CA
3	Demonstration	Belleville, NJ	42	Matched	Concord, CA
3	Matched	Flushing, NY	42	Matched	Downey, CA
3	Matched	North Tarrytown, NY	43	Demonstration	Bakersfield, CA
3	Matched	Valley Stream, NY	43	Matched	Chula Vista, CA
3	Matched	Oceanside, NY	43	Matched	Oxnard, CA
3	Matched	Rockville Centre, NY	43	Matched	Upland, CA
3	Matched	Bay Shore, NY	43	Matched	Visalia, CA
3	Matched	New Rochelle, NY	43	Matched	Palm Springs, CA
3	Matched	Huntington, NY	43	Matched	Turlock, CA
3	Matched	Nyack, NY	43	Matched	Corona, CA
3	Matched	Port Jefferson, NY	43	Matched	Hemet, CA
4	Demonstration	Bridgeton, NJ	44	Demonstration	Loma Linda, CA
4	Matched	Niagara Falls, NY	44	Matched	Fresno, CA
5	Demonstration	Camden, NJ	45	Demonstration	Los Angeles, CA
5	Matched	Flushing, NY	45	Matched	Burbank, CA
5	Matched	Nyack, NY	46	Demonstration	Martinez, CA
5	Matched	Rockville Centre, NY	46	Matched	Monterey Park, CA
5	Matched	Huntington, NY	46	Matched	Gardena, CA
5	Matched	New Rochelle, NY	46	Matched	Garden Grove, CA
5	Matched	Bay Shore, NY	46	Matched	Merced, CA
5	Matched	Valley Stream, NY	46	Matched	Montebello, CA
5	Matched	North Tarrytown, NY	47	Demonstration	Mission Hills, CA
5	Matched	Port Jefferson, NY	47	Matched	Van Nuys, CA
5	Matched	Oceanside, NY	47	Matched	Long Beach, CA
6	Demonstration	Cape May Court House, NJ	47	Matched	Covina, CA
6	Matched	Elmira, NY	47	Matched	Pomona, CA
6	Matched	Plattsburgh, NY	48	Demonstration	Orange, CA
6	Matched	Cortland, NY	48	Matched	Oakland, CA
6	Matched	Newark, NY	48	Matched	Berkeley, CA
7	Demonstration	Dover, NJ	48	Matched	Mission Viejo, CA
7	Matched	Manhattan, NY	48	Matched	Santa Monica, CA
7	Matched	West Islip, NY	49	Demonstration	Riverside, CA
7	Matched	White Plains, NY	49	Matched	Visalia, CA
7	Matched	Brooklyn, NY	49	Matched	Palm Springs, CA
7	Matched	Staten Island, NY	49	Matched	Turlock, CA
7	Matched	Stony Brook, NY	49	Matched	Corona, CA
7	Matched	Mineola, NY	49	Matched	Hemet, CA
7	Matched	Manhasset, NY	49	Matched	Chula Vista, CA
7	Matched	Jamaica, NY	49	Matched	Oxnard, CA
7	Matched	Suffern, NY	49	Matched	Upland, CA
8	Demonstration	Edison, NJ	50	Demonstration	Sacramento, CA
8	Matched	Smithtown, NY	50	Matched	La Mesa, CA
9	Demonstration	Elizabeth, NJ	50	Matched	Modesto, CA

New Jersey/New York			California		
Demonstration/ comparison set index	Demonstration or matched comparison HSA	HSA name	Demonstration/ comparison set Index	Demonstration or matched comparison HSA	HSA name
9	Matched	Staten Island, NY	51	Demonstration	Salinas, CA
9	Matched	Stony Brook, NY	51	Matched	Chula Vista, CA
9	Matched	Mineola, NY	51	Matched	Oxnard, CA
9	Matched	Manhasset, NY	51	Matched	Upland, CA
9	Matched	Jamaica, NY	51	Matched	Visalia, CA
9	Matched	White Plains, NY	51	Matched	Palm Springs, CA
9	Matched	West Islip, NY	51	Matched	Turlock, CA
9	Matched	Manhattan, NY	51	Matched	Corona, CA
9	Matched	Suffern, NY	51	Matched	Hemet, CA
9	Matched	Brooklyn, NY	52	Demonstration	San Diego, CA
10	Demonstration	Englewood, NJ	52	Matched	Corona, CA
10	Matched	Huntington, NY	52	Matched	Chula Vista, CA
10	Matched	New Rochelle, NY	52	Matched	Hemet, CA
10	Matched	North Tarrytown, NY	52	Matched	Upland, CA
10	Matched	Valley Stream, NY	52	Matched	Visalia, CA
10	Matched	Nyack, NY	52	Matched	Palm Springs, CA
10	Matched	Flushing, NY	52	Matched	Oxnard, CA
10	Matched	Oceanside, NY	52	Matched	Turlock, CA
10	Matched	Port Jefferson, NY	53	Demonstration	San Francisco, CA
10	Matched	Rockville Centre, NY	53	Matched	Daly City, CA
10	Matched	Bay Shore, NY	53	Matched	Pasadena, CA
11	Demonstration	Freehold, NJ	53	Matched	Culver City, CA
11	Matched	Port Jefferson, NY	54	Demonstration	San Jose, CA
11	Matched	New Rochelle, NY	54	Matched	Santa Maria, CA
11	Matched	Huntington, NY	54	Matched	Monterey, CA
11	Matched	Flushing, NY	54	Matched	Poway, CA
11	Matched	North Tarrytown, NY	54	Matched	Oceanside, CA
11	Matched	Nyack, NY	54	Matched	Fallbrook, CA
11	Matched	Oceanside, NY	54	Matched	Carmichael, CA
11	Matched	Rockville Centre, NY	54	Matched	Escondido, CA
11	Matched	Bay Shore, NY	54	Matched	Redding, CA
11	Matched	Valley Stream, NY	54	Matched	Mountain View, CA
12	Demonstration	Hackensack, NJ	54	Matched	Thousand Oaks, CA
12	Matched	Manhattan, NY	54	Matched	Santa Barbara, CA
12	Matched	Manhasset, NY	54	Matched	Santa Cruz, CA
12	Matched	West Islip, NY	55	Demonstration	San Mateo, CA
12	Matched	Brooklyn, NY	55	Matched	Daly City, CA
12	Matched	Jamaica, NY	55	Matched	Pasadena, CA
12	Matched	White Plains, NY	55	Matched	Culver City, CA
12	Matched	Stony Brook, NY	56	Demonstration	Stockton, CA
12	Matched	Staten Island, NY	56	Matched	Visalia, CA
12	Matched	Mineola, NY	56	Matched	Palm Springs, CA
12	Matched	Suffern, NY	56	Matched	Turlock, CA
13	Demonstration	Hoboken, NJ	56	Matched	Corona, CA
13	Matched	Patchogue, NY	56	Matched	Hemet, CA
13	Matched	Long Beach, NY	56	Matched	Chula Vista, CA
14	Demonstration	Holmdel, NJ	56	Matched	Oxnard, CA
14	Matched	Dobbs Ferry, NY	56	Matched	Upland, CA
14	Matched	Bethpage, NY	57	Demonstration	Torrance, CA
14	Matched	Bronxville, NY	57	Matched	Santa Ana, CA
15	Demonstration	Jersey City, NJ	58	Demonstration	Ventura, CA
15	Matched	Patchogue, NY	58	Matched	San Bernardino, CA
15	Matched	Long Beach, NY	58	Matched	National City, CA
16	Demonstration	Lakewood, NJ			
16	Matched	Huntington, NY			
16	Matched	Flushing, NY			
16	Matched	Oceanside, NY			
16	Matched	Rockville Centre, NY			
16	Matched	Nyack, NY			
16	Matched	New Rochelle, NY			
16	Matched	Bay Shore, NY			

New Jersey/New York			California		
Demonstration/ comparison set index	Demonstration or matched comparison HSA	HSA name	Demonstration/ comparison set Index	Demonstration or matched comparison HSA	HSA name
16	Matched	North Tarrytown, NY			
16	Matched	Valley Stream, NY			
16	Matched	Port Jefferson, NY			
17	Demonstration	Livingston, NJ			
17	Matched	Smithtown, NY			
18	Demonstration	Long Branch, NJ			
18	Matched	Flushing, NY			
18	Matched	Huntington, NY			
18	Matched	New Rochelle, NY			
18	Matched	North Tarrytown, NY			
18	Matched	Nyack, NY			
18	Matched	Oceanside, NY			
18	Matched	Port Jefferson, NY			
18	Matched	Rockville Centre, NY			
18	Matched	Bay Shore, NY			
18	Matched	Valley Stream, NY			
19	Demonstration	Manahawkin, NJ			
19	Matched	Bethpage, NY			
19	Matched	Bronxville, NY			
19	Matched	Dobbs Ferry, NY			
20	Demonstration	Montclair, NJ			
20	Matched	Plainview, NY			
20	Matched	Glen Cove, NY			
21	Demonstration	Morristown, NJ			
21	Matched	Smithtown, NY			
22	Demonstration	Mount Holly, NJ			
22	Matched	Huntington, NY			
22	Matched	New Rochelle, NY			
22	Matched	Nyack, NY			
22	Matched	Oceanside, NY			
22	Matched	Port Jefferson, NY			
22	Matched	Rockville Centre, NY			
22	Matched	Bay Shore, NY			
22	Matched	Valley Stream, NY			
22	Matched	Flushing, NY			
22	Matched	North Tarrytown, NY			
23	Demonstration	Neptune, NJ			
23	Matched	Jamaica, NY			
23	Matched	Manhasset, NY			
23	Matched	Mineola, NY			
23	Matched	Manhattan, NY			
23	Matched	Staten Island, NY			
23	Matched	Stony Brook, NY			
23	Matched	Suffern, NY			
23	Matched	Brooklyn, NY			
23	Matched	West Islip, NY			
23	Matched	White Plains, NY			
24	Demonstration	New Brunswick, NJ			
24	Matched	Jamaica, NY			
24	Matched	Manhasset, NY			
24	Matched	Mineola, NY			
24	Matched	Manhattan, NY			
24	Matched	Stony Brook, NY			
24	Matched	Suffern, NY			
24	Matched	Brooklyn, NY			
24	Matched	West Islip, NY			
24	Matched	White Plains, NY			
24	Matched	Staten Island, NY			
25	Demonstration	Newark, NJ			
25	Matched	Long Beach, NY			
25	Matched	Patchogue, NY			

New Jersey/New York			California		
Demonstration/ comparison set index	Demonstration or matched comparison HSA	HSA name	Demonstration/ comparison set Index	Demonstration or matched comparison HSA	HSA name
26	Demonstration	Newton, NJ			
26	Matched	Bethpage, NY			
26	Matched	Bronxville, NY			
26	Matched	Dobbs Ferry, NY			
27	Demonstration	North Bergen, NJ			
27	Matched	Mount Kisco, NY			
27	Matched	Mount Vernon, NY			
27	Matched	Peekskill, NY			
27	Matched	Riverhead, NY			
27	Matched	Southampton, NY			
27	Matched	Carmel, NY			
27	Matched	Cooperstown, NY			
28	Demonstration	Passaic, NJ			
28	Matched	Huntington, NY			
28	Matched	New Rochelle, NY			
28	Matched	North Tarrytown, NY			
28	Matched	Nyack, NY			
28	Matched	Oceanside, NY			
28	Matched	Port Jefferson, NY			
28	Matched	Rockville Centre, NY			
28	Matched	Bay Shore, NY			
28	Matched	Valley Stream, NY			
28	Matched	Flushing, NY			
29	Demonstration	Paterson, NJ			
29	Matched	Jamaica, NY			
29	Matched	Manhasset, NY			
29	Matched	Mineola, NY			
29	Matched	Manhattan, NY			
29	Matched	Staten Island, NY			
29	Matched	Stony Brook, NY			
29	Matched	Suffern, NY			
29	Matched	Brooklyn, NY			
29	Matched	West Islip, NY			
29	Matched	White Plains, NY			
30	Demonstration	Phillipsburg, NJ			
30	Matched	Rhinebeck, NY			
30	Matched	Warwick, NY			
31	Demonstration	Point Pleasant, NJ			
31	Matched	Plainview, NY			
31	Matched	Glen Cove, NY			
32	Demonstration	Pompton Plains, NJ			
32	Matched	Plainview, NY			
32	Matched	Glen Cove, NY			
33	Demonstration	Princeton, NJ			
33	Matched	Smithtown, NY			
34	Demonstration	Red Bank, NJ			
34	Matched	Huntington, NY			
34	Matched	New Rochelle, NY			
34	Matched	North Tarrytown, NY			
34	Matched	Nyack, NY			
34	Matched	Oceanside, NY			
34	Matched	Port Jefferson, NY			
34	Matched	Rockville Centre, NY			
34	Matched	Bay Shore, NY			
34	Matched	Valley Stream, NY			
34	Matched	Flushing, NY			
35	Demonstration	Secaucus, NJ			
35	Matched	Bethpage, NY			
35	Matched	Bronxville, NY			
35	Matched	Dobbs Ferry, NY			
36	Demonstration	Somerville, NJ			

New Jersey/New York			California		
Demonstration/ comparison set index	Demonstration or matched comparison HSA	HSA name	Demonstration/ comparison set Index	Demonstration or matched comparison HSA	HSA name
36	Matched	Plainview, NY			
36	Matched	Glen Cove, NY			
37	Demonstration	Summit, NJ			
37	Matched	Smithtown, NY			
38	Demonstration	Toms River, NJ			
38	Matched	Smithtown, NY			
39	Demonstration	Trenton, NJ			
39	Matched	Hornell, NY			
39	Matched	Ithaca, NY			
39	Matched	Jamestown, NY			
39	Matched	Lewiston, NY			
39	Matched	Lowville, NY			
39	Matched	Middletown, NY			
39	Matched	Newburgh, NY			
39	Matched	Oneida, NY			
39	Matched	Oneonta, NY			
39	Matched	Port Jervis, NY			
39	Matched	Poughkeepsie, NY			
39	Matched	Amsterdam, NY			
39	Matched	Saranac Lake, NY			
39	Matched	Schenectady, NY			
39	Matched	Syracuse, NY			
39	Matched	Binghamton, NY			
39	Matched	Troy, NY			
39	Matched	Utica, NY			
39	Matched	Watertown, NY			
39	Matched	Catskill, NY			
39	Matched	Clifton Springs, NY			
39	Matched	Glens Falls, NY			
40	Demonstration	Willingboro, NJ			
40	Matched	Mount Kisco, NY			
40	Matched	Mount Vernon, NY			
40	Matched	Peekskill, NY			
40	Matched	Riverhead, NY			
40	Matched	Southampton, NY			
40	Matched	Carmel, NY			
40	Matched	Cooperstown, NY			
41	Demonstration	Woodbury, NJ			
41	Matched	Plainview, NY			
41	Matched	Glen Cove, NY			

Table B.4. 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
New Jersey	2012	2014	Through 2014	4	Yes
Texas	2011	2011	Through 2013[b]	9	Yes
New York	2014	2015	Through 2014[c]	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] 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.

[c] 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.

DSRIP = Delivery System Reform Incentive Payment.

SECTION B.2. ESTIMATING DEMONSTRATION EFFECTS

A. California and New Jersey: difference-in-differences

In California and New Jersey, we used a difference-in-differences approach to test the causal effects of DSRIP demonstrations on patient outcomes. This econometric technique is used to determine whether outcomes have changed differently for the demonstration group than for the comparison group after implementation of the intervention of interest. To be specific, the patient-level models estimated the regression-adjusted difference in the change (from pre to post-intervention period) in the average level of outcomes between patients residing in DSRIP HSAs and patients residing in similar non-DSRIP HSAs.

For each state, we estimated a patient-level regression model for an outcome (*y)* for person (*i*), in community (*j*)[22] at time (*t):*

$$(B.1) \qquad y_{ijt} = \beta_0 + \beta_1 Post_{ijt} + \beta_2 Demo_{ij} + \beta_3 Demo_{ij} * Post_{ijt} + \beta_4 W_{ij} + \varepsilon_{ijt}$$

This model includes three types of covariates:

- $Post_{ijt}$ is an intervention indicator, equal to one if the observation is in the post-period, and equal to zero if the observation is in the pre-period
- $Demo_{ij}$ is a demonstration indicator, equal to one if community *j* is affected by DSRIP and equal to zero if community *j* is in the comparison group
- W_{ij} contains patient-level characteristics measured at baseline, such as age, gender, and presence of chronic conditions, as well as hospital and community-level characteristics, such as a community's number of hospital beds and hospitals and its residents' median income

Because the three clinical outcomes of interest—(1) whether a patient had an ED visit during a given quarter; (2) whether a patient had a follow-up after an ED discharge; and (3) whether a diabetes patient had a hemoglobin A1C test in a given year—are binary, we used a linear probability regression model to estimate the impacts of the DSRIP program on the outcomes. A linear probability model relies on a linear regression to estimate the probability of experiencing a positive outcome. As a result, the parameter of interest in this model, β_3, could be interpreted as the regression-adjusted difference in the probability, from pre to post-intervention period, in the average level of outcomes between patients residing in DSRIP HSAs and patients residing in similar non-DSRIP HSAs. A positive value for β_3 indicates that the average level of outcome increased by a larger percentage or decreased by a smaller percentage in the DSRIP HSAs than in similar non-DSRIP HSAs from the pre- to post-intervention period.

[22] The community is defined by the HSA.

As a sensitivity test, we also estimated logit models and found that results were consistent with those from the linear probability modes.

B. Texas: simple interrupted time series

Because Texas implemented the DSRIP program in the entire state and because none of our potential comparison states were similar to Texas, our interim evaluation did not involve a comparison group. Our analytic strategy relied on implementing a simple interrupted time series design to estimate whether the level or trends in the outcomes of interest in the post-intervention period were significantly different than what would be expected in the absence of the intervention. A simple interrupted time series design relies on an assumption that the trajectory of the outcome in the pre-intervention period can predict the expected trajectory in the post-intervention period.

As in the other states, we estimated a patient-level linear probability regression model for an outcome (*y)* for person (*i*), in community (*j*)[23] at time (*t)* but used the following equation.

$$(B.2) \qquad y_{ijt} = \beta_0 + \beta_1 Post_{ijt} + \beta_2 time_{ijt} + \beta_3 time_{ijt} * Post_{ijt} + \beta_4 W_{ij} + \varepsilon_{ijt} .$$

In this model, $time_{ijt}$ is time (measured in quarters) since the DSRIP program was implemented in community *j* at time *t*, with $time_{ijt} = 0$ in the first quarter after DSRIP implementation.

There are two parameters of interest in this model: change in the level (β_1) and change in the trend (β_3) in the outcome of interest. A positive value for (β_1) indicates that we observed an increase in the level of the outcome of interest with implementation of the DSRIP program. A positive value for β_3 indicates that there was a gradual (quarterly or annual) increase in the trend of the outcome of interest after the implementation of the DSRIP program.

Note: cross-sectional analysis

In the DSRIP design supplement (Baller et al. 2017), we proposed to conduct a sensitivity test using cross-sectional analyses, in which we would aim to reinforce the credibility of our estimated demonstration effects by comparing outcomes in communities with higher and lower levels of DSRIP funding. Such estimates would not rely on the assumption that demonstration and comparison communities are comparable, which is required for differences-in-differences estimates, nor on the assumption that baseline levels and trends of outcome variables persist into the post-period, which is relevant to both difference-in-differences and interrupted time series models. However, cross-sectional analysis would only be valid if the level of funding were independent of the outcome variables, and, upon review of project documents, we concluded that this assumption was not valid for the DSRIP program.

[23] The community is defined by the HSA.

APPENDIX C

REGRESSION RESULTS

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

Table C.1. Impact of DSRIP in California between 2009 and 2014

	ED visits[a]	Follow-up after an ED visit for ambulatory care sensitive condition[b]	Diabetes testing[b]
Sample size (N)	25,960,740	98,960	305,462
Demonstration (versus comparison) (percentage point difference)	−0.4**	1.7**	−1.1*
Post-period (versus pre-period) (percentage point difference)	−0.7**	−1.5**	−7.7**
Post*demonstration (percentage point difference)	0.4**	−0.8	1.7**

Source: Mathematica analysis of MAX and Alpha-MAX data from California, 2009 to 2014.

Note: Each column represents results from a separate difference-in-differences analysis using linear probability regression models. Post-period is a binary indicator variable, which equals one in the years after DSRIP was implemented. Demonstration is a binary indicator variable, which equals one for beneficiaries living in California HSAs subject to DSRIP, and zero otherwise. The Post*demonstration interaction term is the main difference-in-differences effect, or the difference between the comparison and demonstration groups in their probability of experiencing the outcome before and after DSRIP implementation. ED visits and follow-up after an ED visit for ambulatory care sensitive conditions are measured at the person-quarter level. Diabetes testing is measured at the person-year level. The models also controlled for sex, age, clinical conditions, median household income, beds per resident, and hospitals per HSA.

[a] A positive coefficient on the post*demonstration term indicates a worse outcome.

[b] A positive coefficient on the post*demonstration term indicates a better outcome.

*indicates statistical significance less than 0.01 but greater than 0.001.

**indicates statistical significance less than 0.001.

ED = emergency department; HSA = hospital service area.

Table C.2. Impact of DSRIP in New Jersey between 2011 and 2014

	ED visits[a]	Follow-up after an ED visit for ambulatory care sensitive condition[b]	Diabetes testing[b]
Sample size (N)	16,225,280	154,144	253,865
Demonstration (versus comparison) (percentage point difference)	5.4**	−2.8**	2.2**
Post-period (versus pre-period) (percentage point difference)	−1.7**	-0.2	−2.4**
Post*demonstration (percentage point difference)	0.2**	0.6	8.0**

Source: Mathematica analysis of MAX and Alpha-MAX data from New Jersey and New York, 2011 to 2014.

Note: Each column represents results from a separate difference-in-differences analysis using linear probability regression models. Post-period is a binary indicator variable, which equals one in the years after DSRIP was implemented. Demonstration is a binary indicator variable, which equals one for beneficiaries living in New Jersey HSAs subject to DSRIP and zero for beneficiaries living in comparison New York HSAs. The Post*demonstration interaction term is the main difference-in-differences effect, or the difference in the probability of experiencing the outcome between the comparison and demonstration groups, before and after DSRIP implementation. ED visits and follow-up after an ED visit for an ambulatory care sensitive condition are measured at the person-quarter level. Diabetes testing is measured at the person-year level. The models also controlled for sex, age, clinical conditions, median household income, beds per resident, and hospitals per HSA.

[a] A positive coefficient on the post*demonstration term indicates a worse outcome.

[b] A positive coefficient on the post*demonstration term indicates a better outcome.

*indicates statistical significance less than 0.01 but greater than 0.001.
**indicates statistical significance less than 0.001.

ED = emergency department; HSA = hospital service area.

Table C.3. Impact of DSRIP in Texas between 2009 and 2013

	ED visits[a]	Follow-up after an ED visit for ambulatory care sensitive condition[b]	Diabetes testing[b]
Sample size (N)	2,740,880	23,102	22,870
Time (percentage point difference)	0.0**	-0.1	-1.6*
Post-period (versus pre-period) (percentage point difference)	0.0	-2.6	-3.8*
Time*'Post-period (percentage point difference)	0.1**	0.3	5.6**

Source: Mathematica analysis of MAX and Alpha-MAX data from Texas, 2009 to 2013

Note: Each column represents results from a separate simple interrupted time series using linear probability regression models. Post-period is a binary indicator variable that equals one in the years after DSRIP was implemented. This is a measure of the change in level of the outcome of interest at the point in which DSRIP is implemented. Time is measured in quarters for the measures of ED visits and follow-up after an ED visit for an ambulatory care sensitive condition, and it is measured annually for the diabetes testing measure. The interaction between Time and Post-period estimates the change in the trend of the outcome of interest in the post-period. ED visits and follow-up after an ED visit for an ambulatory-sensitive condition are measured at the person-quarter level. Diabetes testing is measured at the person-year level. The model also controlled for sex, age, clinical conditions, median household income, beds per resident, and hospitals per HSA.

[a] A positive coefficient on the post*time term indicates a worse outcome.

[b] A positive coefficient on the post*time term indicates a better outcome.

*indicates statistical significance less than 0.01 but greater than 0.001

**indicates statistical significance less than 0.001

ED = emergency department; HSA = hospital service area.

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

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

www.mathematica-mpr.com

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

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA Policy Research

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