Mathematica
Progress Together

MEDICAID 1115 DEMONSTRATIONS

Medicaid Section 1115 Demonstrations Summative Evaluation Report

Delivery System Reform Incentive Payments

August 2020

Julia Baller

Heather Dahlen

Natasha Reese-McLaughlin

Natalya Verbitsky-Savitz

Christina Ribar

Aaron Swaney

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

ACKNOWLEDGMENTS

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

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

Medicaid Delivery System Reform Incentive Payment (DSRIP) demonstrations, implemented under section 1115 demonstration authority, encourage hospitals and other health care providers 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 payment. Although DSRIP demonstrations share the same broad goals and operational framework, they vary considerably across states and have evolved over time.

This is the summative evaluation of four DSRIP section 1115 demonstrations in California, New Jersey[1], New York and Texas. The report addresses four research questions about the overall effect of the DSRIP demonstration on key outcomes related to transforming the delivery system and clinical processes in each state:

1. What is the overall effect of DSRIP demonstrations on shifting care away from emergency department (ED) and inpatient settings? To address this question, we examined changes in ED visits, avoidable ED visits, and hospital discharges for ambulatory care sensitive conditions (ACSCs).

2. What is the overall effect of DSRIP demonstrations on use of primary care and preventive services? To address this question, we examined changes in ambulatory care visits for adults and primary care visits for children and adolescents.

3. What is the overall effect of DSRIP demonstrations on use of behavioral health services? To address this question, we examined changes in behavioral health visits.

4. What is the overall effect of DSRIP demonstrations on clinical care processes? To address this question, we examined changes in hemoglobin A1c (HbA1c) testing among beneficiaries with diabetes and follow-up within seven days after an ED visit for ACSCs.

Given substantial differences among demonstration states in the design and scope of their DSRIP demonstrations and their Medicaid programs, we conducted state-specific analyses. 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 or community-level characteristics. In California, where the demonstration programs were implemented in some parts of the state, we relied on a difference-in-differences approach that compared outcomes before and after demonstration implementation for Medicaid beneficiaries living in California communities affected by DSRIP to beneficiaries living in similar communities not affected by DSRIP in California, Virginia, and Washington. In New Jersey, New York, and Texas, where the demonstration programs were implemented in all or most communities throughout the state, we

---

[1] Due to severe data quality concerns with the New Jersey Medicaid administrative data, we were unable to construct outcomes measures using this data source. The evaluation of the New Jersey DSRIP demonstration is limited to outcomes constructed using the Healthcare Cost and Utilization Project hospital discharge data.

relied on a simple interrupted time series approach in which we examined changes in both the level and trend for outcomes of interest before and after the demonstration was implemented.

Analyses in California, New York, and Texas relied on Medicaid administrative claims data from 2009 to 2017. We combined three Medicaid administrative data sources: the Medicaid Analytic eXtract (MAX), the early version of MAX known as Alpha-MAX, and the T-MSIS Analytic Files (TAFs). Additional analysis in New Jersey and Texas relied on state Healthcare Cost and Utilization Project (HCUP) data, the largest collection of longitudinal hospital care data in the United States. In contrast to Medicaid administrative data, HCUP data include all people who receive care, including those without insurance—a key target population for DSRIP demonstrations. We also used data from the American Community Survey, the Dartmouth Atlas of Health Care, and health professional shortage areas to construct covariates included in the regression analyses.

Table ES.1 below reveals the main effect of DSRIP for each state on the outcomes of interest. Although the results were mixed, we did see some promising findings:

- In California, results were generally unfavorable or non-significant. Both the initial DSRIP and its successor program, the Public Hospital Redesign and Incentives in Medi-Cal (PRIME), were associated with unfavorable ED outcomes—ED visits and avoidable ED visits increased after DSRIP and PRIME implementation, relative to the comparison group. We found no impacts on measures of use of primary and preventive care. DSRIP was not associated with changes in behavioral health service use, but PRIME was associated with an increase, a favorable outcome. Results relating to clinical care processes were mixed— neither program was associated with HbA1c testing among beneficiaries with diabetes. DSRIP was associated with a decrease in follow-up after an ED visit for ACSCs, but PRIME was not associated with this measure.

- In New Jersey, DSRIP was associated with a favorable outcome relating to shifting care away from inpatient settings—hospital discharges for ACSCs decreased after DSRIP implementation, but the magnitude was small.

- In New York, DSRIP was associated with favorable outcomes relating to shifting care away from the ED and use of ambulatory care for adults, though it was not associated with changes in primary care visits in children and adolescents. However, it was associated with an unfavorable decline in use of behavioral health services and was not associated with measures of clinical care processes.

- In Texas, results were mixed. DSRIP was associated with favorable outcomes relating to ED visits, avoidable ED visits, ambulatory care visits for adults, behavioral health visits, and follow-up within seven days after an ED visit for ACSCs. However, it was associated with unfavorable outcomes relating to hospital discharges for ACSCs, primary care visits for children and adolescents, and HbA1c testing among beneficiaries with diabetes.

Table ES.1. Estimated impact of DSRIP: summary

	California (DSRIP)	California (PRIME)	New Jersey	New York	Texas
	Relative change in post-period—DSRIP group	Relative change in post-period—DSRIP group	Change in trend in post-period	Change in trend in post-period	Change in trend in post-period
ED visits	—	—	n.a.	+	+
Avoidable ED visits	—	—	n.a.	+	+
Hospital discharges for ACSCs	n.a.	n.a.	+	n.a.	—
Ambulatory care visits for adults	NS	NS	n.a.	+	+
Primary care visits for children and adolescents	NS	NS	n.a.	NS	—
Behavioral health visits	NS	+	n.a.	—	+
HbA1c testing	NS	NS	n.a.	NS	—
Follow-up after an ED visit for ACSC	—	NS	n.a.	NS	+

Source: Mathematica analysis of Medicaid administrative data and HCUP data.

Note: For ED visits, avoidable ED visits, and hospital discharges for ACSCs, an increase in the outcome of interest signifies a negative outcome whereas a decrease signifies a positive outcome.

+ signifies a favorable outcome; — signifies an unfavorable outcome; NS = not significant; n.a. = not applicable

ACSC = ambulatory care sensitive conditions; DSRIP = delivery system reform incentive payment; ED = emergency department; PRIME = Public Hospital Redesign and Incentives in Medi-Cal.

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

I. INTRODUCTION

Delivery System Reform Incentive Payment (DSRIP) demonstrations, which are authorized and implemented under Medicaid section 1115 authority, encourage hospitals and other health care providers 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 payment.[2]

In 2014, the Center for Medicaid and CHIP Services within the Centers for Medicare & Medicaid Services (CMS) contracted with Mathematica to conduct an independent national evaluation of the implementation and outcomes of certain Medicaid section 1115 demonstrations. The purpose of this cross-state evaluation is to help policymakers at the state and federal levels understand the extent to which innovations further the goals of the Medicaid program, as well as to inform CMS decisions regarding future section 1115 demonstration approvals, renewals, and amendments.

This is the summative evaluation of the first four DSRIP demonstration programs in California, New Jersey, New York, and Texas. It builds on an interim evaluation published in 2018 (Baller 2018) and is based on an updated evaluation design published in 2019 (Baller 2019). These four were the only states in which the program had at least one year of data available in the post-period.[3]

The DSRIP demonstrations share the same broad goals and operational framework, as shown in the logic model developed for this evaluation (Figure I.1).[4] However, they vary considerably in other respects across the study states and have evolved over time. (Table I.1 provides DSRIP demonstration characteristics by state.) For example, in California and New Jersey, only hospital systems are eligible for DSRIP incentive payments, whereas DSRIP demonstrations in New York and Texas have more expansive provider eligibility criteria and require participating providers to form regional collaborations. Further, some states prioritized certain outcomes over others. For instance, New York's program explicitly aimed to reduce avoidable hospital use by 25 percent, whereas other states did not target this measure. Finally, there is considerable within-state variation across providers in terms of the number and types of projects they are implementing and the number and types of milestones and measures they report.

---

[2] Preparing for value-based payment or participating in alternative payment models by contracting with managed care plans is an explicit goal of DSRIP demonstrations in California (Public Hospital Redesign and Incentives in Medi-Cal), Massachusetts (Delivery System Transformation Initiatives and DSRIP), New Hampshire, New York, and Washington.

[3] As of June 2019, eight states—California, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island, Texas, and Washington—had active DSRIP demonstrations. Massachusetts and New Hampshire did not begin implementation until 2017, the most recent year of data available for the summative evaluation, and Rhode Island and Washington did not begin implementation until 2018. Massachusetts implemented a precursor to its DSRIP demonstration, the Delivery System Transformation Initiatives, from 2012 through 2017. Although certain attributes of the program resemble other DSRIP demonstrations, the state does not consider it a DSRIP demonstration. As a result, the analysis excluded these states.

[4] To develop the logic model, we reviewed state documentation, including Special Terms and Conditions and relevant attachments. States articulated the pathways depicted in the logic model in this documentation, with a focus on strategies and outcomes.

Figure I.1. DSRIP demonstration: Logic model.
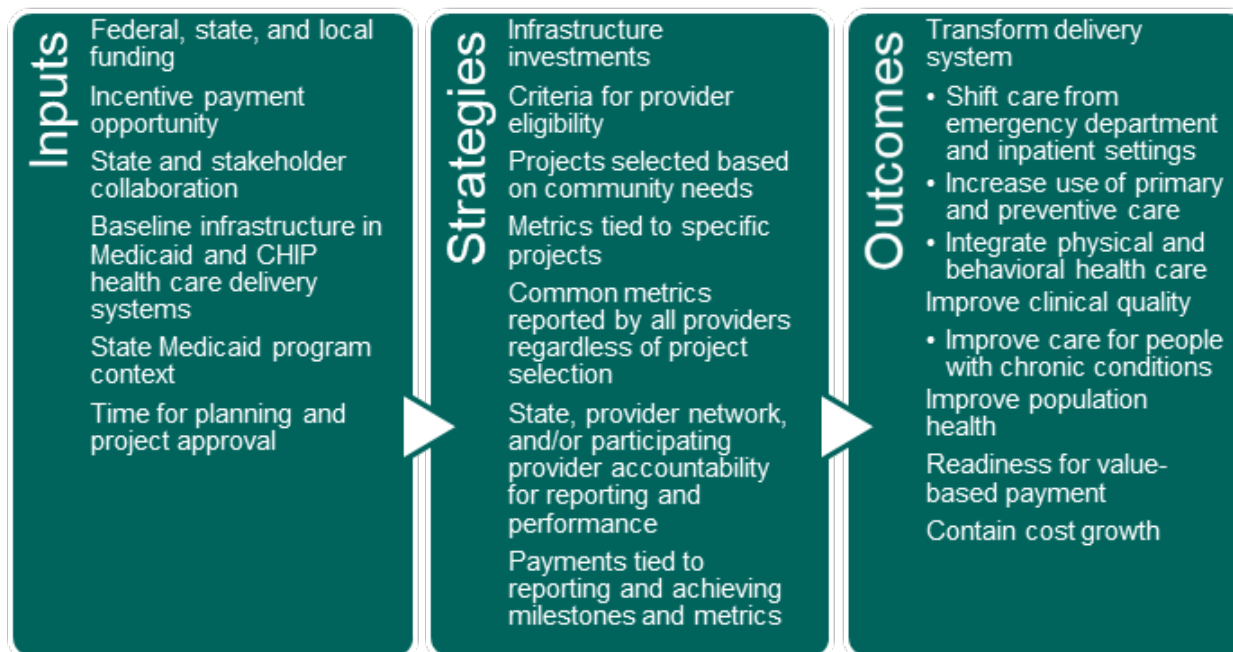
Table I.1. DSRIP demonstration characteristics in four states

Characteristic	California--DSRIP	California--PRIME	New Jersey[a]	New York	Texas[a]
Approval date	11/1/2010	12/30/2015	10/1/2012	4/14/2014	12/12/2011
Expiration date	12/31/2015	12/31/2020	6/30/2020	3/31/2020	9/20/2021
Total program funding	$6.671B	$7.464B	$1.083B	$6.417B	$26.118B
Additional funding introduced by DSRIP?	Yes	Yes	No	Yes	Yes
Type of providers eligible to receive incentive payments	Designated Public Hospital Systems	Designated Public Hospital Systems and District Municipal Public Hospitals	Acute care hospitals	Providers in Performing Provider systems	Regional consortia of providers
Number of providers participating as of August 2020	17 Designated Public Hospital Systems	17 Designated Public Hospital Systems and 34 District Municipal Public Hospitals	49 hospitals	91,603 providers	335 providers in 20 Regional Healthcare Partnerships
Number of projects	221	269	49	259	1,450[b]

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 & Medicaid Services and California Health and Human Services Agency. "Special Terms and Conditions, No. 11-W-00193/9, California Medi-Cal 2020 Demonstration." Approved December 31, 2015 through December 31, 2020.

U.S. Centers for Medicare & 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.

Table I.1 (*continued*)

> U.S. Centers for Medicare & 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 New York State Department of Health. "Special Terms and Conditions, No. 11-W-00114/2, New York Partnership Plan Section 1115 Medicaid Demonstration." Approval Period: August 1, 2011 – December 31, 2014; as amended April 14, 2014.
>
> U.S. 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." Approval period: December 12, 2011, through September 30, 2016; amendment approved February 26, 2015.
>
> U.S. 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." Approval period: January 2, 2018, through September 30, 2022; approved December 21, 2017.

[a] Program currently in a renewal period.

[b] Number of projects in the first demonstration period. State is not implementing projects in the second demonstration period.

DSRIP = delivery system reform incentive payment; PRIME = Public Hospital Redesign in Medi-CAL.

The summative evaluation addresses four research questions, developed in consultation with CMS about the overall effect of DSRIP demonstrations on key outcomes related to transforming delivery systems and clinical processes in each state (from a federal perspective):

1. What is the overall effect of DSRIP demonstrations on shifting care away from emergency department (ED) and inpatient settings?

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

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

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

To address these questions, the evaluation focuses on eight clinical process measures that (1) reflect the DSRIP demonstrations' overall goal of transforming the delivery system, as characterized by an increased use of primary and behavioral health care and improved clinical care; and (2) are likely to respond relatively quickly to DSRIP initiatives. These outcomes include the following:

- Emergency department (ED) visits

- Avoidable ED visits

- Hospital discharges for ambulatory care sensitive conditions (ACSCs) (Prevention quality indicators [PQI] composite of chronic conditions)

- Adult ambulatory care visits

- Child and adolescent access to primary care practitioners

- Behavioral health visits

- Hemoglobin A1C testing for patients with diabetes

- Follow-up after an ED visit for ACSCs (including asthma, chronic obstructive pulmonary disease, hypertension, or diabetes)

We calculated outcomes using two main data sources: (1) Medicaid administrative data and (2) the Healthcare Cost and Utilization Project (HCUP) state inpatient databases funded by the Agency for Healthcare Research and Quality (AHRQ). Table I.2 provides a brief overview of the evaluation design.

Table I.2. Overview of design for the summative evaluation

State	Available data sources	Model	Pre-period	Post-period(s)[a]
California	Medicaid administrative data	Difference-in-differences	2009-2011	2012-2015 2015-2017
New Jersey	Healthcare Cost and Utilization Project data	Simple interrupted time series	2009-2013	2014-2015
New York	Medicaid administrative data	Simple interrupted time series	2009-2014	2015-2017
Texas	Medicaid administrative data Healthcare Cost and Utilization Project data	Simple interrupted time series	2009-2011	2012-2017

[a] The post-period for California incorporates the initial DSRIP period and the follow-on PRIME program. Although New Jersey and Texas are in their respective second demonstration periods, the data in the post-period cover only the initial DSRIP demonstrations.

The rest of this section describes DSRIP demonstrations in California, New Jersey, New York, and Texas in detail. We then present methods (Chapter II), results (Chapter III), and conclude with a discussion (Chapter IV).

A. California

Shortly after the passage of the Patient Protection and Affordable Care Act (ACA) in 2010, California began implementing its section 1115 demonstration, Bridge to Reform, which included the nation's first DSRIP demonstration. Through DSRIP, the state sought to build capacity for treating a newly insured population and improve care for Medicaid beneficiaries and the uninsured. Implementation of the state's DSRIP demonstration began in November 2010 and concluded in October 2015.

Only designated public hospital (DPH) systems were eligible for DSRIP, and 17 DPHs, consisting of 21 hospitals, participated in the program.[5] Over the course of the demonstration, California and CMS made $6.67 billion in federal and nonfederal funding 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 proportion of total funding available to each DPH was

---

[5] Six DPHs are multihospital systems, leading to 21 total hospitals.

determined based on the DPH system's cost, number of low-income individuals served, differences in system infrastructure, and differences in patient populations. Although the DSRIP demonstration in California aimed to improve the five areas listed above, programs used most of the funding for infrastructure investments (Lane et al. 2019).

California is now implementing its successor DSRIP demonstration—the Public Hospital Redesign and Incentives in Medi-Cal (PRIME) program, which makes $7.76 billion available in combined funding over the five-year demonstration period (2015–2020). This program aims to leverage the foundational infrastructure improvements made during the original DSRIP demonstration by establishing more ambitious performance targets and designing projects and associated metrics to drive systemwide changes.

Under PRIME, the state expanded eligibility to 38 district and municipal hospitals (DMPHs) in addition to the 17 DPHs. Today, a total of 51 entities are participating (four DMPHs have dropped out of the demonstration as of August 2020) and eligible to receive funding. Participants are implementing projects that fall into three domains: (1) outpatient delivery system transformation and implementation, (2) targeted high-risk or high-cost populations, and (3) resource utilization efficiency. Further, all participating entities are expected to show improvement on performance metrics, and DPHs are expected to demonstrate progress toward adopting risk-based alternative payment models (APMs) with managed care plans.

B. New Jersey

Originally authorized in 2012 and extended in 2017, implementation of the New Jersey DSRIP demonstration began in 2013 and is approved to continue through 2020. The demonstration aims to improve access, quality of care, and health outcomes among the state's low-income population (State of New Jersey Department of Health 2013). Initially, the DSRIP demonstration replaced the state's Hospital Relief Subsidy Fund, which paid 55 acute care hospitals to offset uncompensated care costs based on the amount of care delivered to Medicaid beneficiaries and the uninsured. With total funding starting at $583 million, the initial DSRIP demonstration provided no additional funding beyond what was previously available through the subsidy fund to the 49 hospitals that opted to participate. However, by tying this funding to implementing projects and reporting and improving upon measures, CMS and the state expected improvements in clinical care and population health.

When the demonstration began in 2013, each hospital selected a single project to implement over the course of the five-year demonstration, and those opting to participate in the extension period continued implementing the selected project. Projects focus on one of the following chronic conditions: asthma, behavioral health, cardiac care, chemical addiction and substance use, diabetes, HIV/AIDS, obesity, or pneumonia.

To receive incentive payments in the initial demonstration period, hospitals had to achieve specific milestones across four stages: (1) infrastructure development, (2) chronic medical condition redesign and management, (3) quality improvements, and (4) population-focused

improvements. Hospitals also had to report a set of project-specific metrics (Stage 3 measures) and universal metrics (Stage 4 measures).

When CMS approved an extension to the demonstration in 2017, it made an additional $499.8 million available to hospitals opting to participate in the extension period. After a transition year between the initial DSRIP period and the extension period in 2018, the state moved away from milestone reporting and placed greater emphasis on pay-for-performance. Hospitals now must demonstrate improvement on at least 50 percent of project-specific Stage 3 measures and they have to improve performance on universal Stage 4 metrics.

C. New York

Approved in April 2014, New York's DSRIP demonstration seeks to (1) transform the state's safety net system for Medicaid beneficiaries and low-income uninsured populations, (2) reduce avoidable hospital use and improve performance on quality and population health measures across the state, and (3) sustain the transformation through DSRIP by nudging Medicaid managed care to value-based purchasing (VBP) models. The demonstration makes $6.4 billion in combined federal and nonfederal funding available over a six-year implementation period (including a pre-planning year and five years of implementation from April 2015 through December 2019).

To achieve these goals, the state requires safety net providers to form coalitions, known as Performing Provider Systems (PPSs). Each PPS must form a governing body consisting of both health care and social services providers that oversees planning and implementing DSRIP projects, as well as distributing DSRIP funds. Further, they are accountable to the state for meeting certain benchmarks. PPSs select delivery reform projects from three domains: (1) system transformation (for example, creating patient-centered medical homes and improving care transitions from the hospital); (2) clinical improvement, especially integrating primary and behavioral health services; and (3) population-wide health promotion, which focuses on improving health outcomes for patient groups with special needs (such as pregnant women and people with HIV/AIDS). The 25 PPSs are implementing a total of 258 projects, with 14 PPSs implementing a special project on engaging the uninsured and low-use Medicaid populations and linking them to primary and preventive services.

Funding for the New York DSRIP demonstration is tied to performance on metrics, both for individual PPSs and the state. A combination of the number of attributed Medicaid and uninsured lives along with the number of DSRIP projects being implemented determines each PPS's initial funding amount. Over the course of the demonstration, performance expectations ramp up, and the percentage of funding tied to performance increases across all domains by the final demonstration year. PPSs can receive additional funding from supplemental pools that also depend on their metric performance. At the state level, a portion of federal DSRIP funding is at risk based on statewide performance on four milestones starting in the third demonstration year: (1) delivery system improvement; (2) project-specific and population-wide performance; (3) Medicaid spending cost growth containment; and (4) converting Medicaid managed care

payments to VBP models. Five, 10, and 20 percent of funding is at risk based on statewide performance in Demonstration Years 3, 4, and 5, respectively.

D. Texas

CMS approved the initial Texas DSRIP demonstration in December 2011, an extension through 2017, and a second demonstration period from 2017 through 2021. The demonstrations aim to improve access to care, quality of care, and health outcomes, and to reduce costs by transforming the delivery of care to Medicaid beneficiaries and the uninsured.

The initial demonstration made a total of $11.4 billion available to hospitals and other providers (CMS 2015). To encourage regional coordination, the state required providers to organize into regional networks, known as regional health care partnerships (RHPs). A public hospital or governmental entity leads each of the 20 RHPs; the hospital or entity assumes responsibility for coordinating DSRIP activities. To determine funding allocations for each RHP, the state considered, as of 2011, the percentage of the state's population living in poverty in the region, the percentage of Medicaid acute care payments made in the region, and the percentage of total supplemental payments made to the region.

In the initial demonstration, the 335 participating providers (including hospitals, counties, community health centers, and other provider types) within each RHP selected projects and reported metrics. Providers selected projects from four categories: (1) infrastructure development, (2) program innovation and redesign, (3) quality improvements, and (4) population-based improvements, and they implemented a total of 1,451 projects across the state. The providers earned payments primarily through pay-for-reporting, but some projects included pay-for-performance metrics in later years of the demonstration.

In the second demonstration, CMS made an additional $14.7 billion dollars available to providers. Texas fundamentally changed its demonstration design, reflecting an evolution from implementing projects and reporting project-level milestones and metrics to reporting and improving performance on measure bundles. The state now evaluates providers on four categories of reporting: (1) qualitative reporting of core activities, including APM arrangements and collaborative activities; (2) the number of Medicaid and low-income or uninsured patients served by each performing provider; (3) quantitative reporting of measures or measure bundles, depending on the type of provider; and (4) statewide reporting of measure bundles.

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

II. METHODS

In this chapter, we provide an overview of the research questions and outcome measures included in the evaluation, the main data sources, the population of beneficiaries included in our study sample, and the analytic approaches used to estimate the demonstration effects. Detailed technical descriptions are available in Appendix A, which contains additional information on our strategy for matching demonstration and comparison communities (Section A.1), Medicaid administrative claims data availability and quality (Tables A.1 and A.2), and the approach to estimating demonstration effects (Section A.2).

A. Research questions and outcome measures

The summative evaluation addresses four research questions about the overall effect of DSRIP demonstrations on key outcomes related to transforming the delivery system and clinical processes in each state:

1. What is the overall effect of DSRIP demonstrations on shifting care away from ED and inpatient settings?

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

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

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

We selected outcome measures for the summative evaluation that address these questions, that DSRIP demonstrations are most likely to influence, and that use administrative data. The measures reflect CMS and state priorities for each DSRIP demonstration and include endorsed measures, measures relevant to the most common clinical focus areas of the projects,[6] and measures in state DSRIP demonstration evaluations, when possible. In addition, our qualitative findings and reviews of state evaluations suggest that the demonstrations might not have observable impacts on health outcomes immediately (Baller et al. 2017a). Therefore, we focused on the most immediate domains of delivery system transformation and clinical processes, rather than on longer-term changes in health outcomes. In Table II.1, we describe each outcome measure we used to address each research question. Following Table II.1, we discuss the outcomes for each research question in more detail.

---

[6] As part of the interim outcomes evaluation, we developed a streamlined, comprehensive taxonomy of clinical focus areas that reflected the key goals of the DSRIP demonstrations to better understand state and provider clinical priorities. We mapped each project to one or more of the clinical focus areas. These areas included appropriate care in appropriate settings, primary care, behavioral health care, diabetes care, cardiac care, care transitions, and care coordination.

Table II.1. Summary of outcome measures and data sources

	Measure description	Hypothesis	Data sources	Population[a]	States with available data
Research question 1: What is the overall effect of DSRIP demonstrations on shifting care away from ED and inpatient settings?					
ED visits[b]	Quarterly count of ED visits that did not result in an inpatient admission per 1,000 beneficiaries	The number of ED visits per 1,000 beneficiaries will decline after DSRIP implementation for demonstration HSAs relative to comparison HSAs (when relevant).	Medicaid administrative data	Medicaid and CHIP beneficiaries	California New York Texas
Avoidable ED visits[b,c]	Quarterly count of avoidable ED visits per 1,000 beneficiaries	Avoidable ED visits per 1,000 beneficiaries will decline after DSRIP implementation for demonstration HSAs relative to comparison HSAs (when relevant).	Medicaid administrative data	Medicaid and CHIP beneficiaries	California New York Texas
Hospital discharges for one of several chronic conditions[b] (PQI 92)	Quarterly count of hospital discharges per 100,000 Medicaid beneficiaries and uninsured for: • Diabetes short-term complications • Diabetes long-term complications • COPD or asthma (in older adults) • Hypertension • Heart failure • Angina without procedure • Uncontrolled diabetes • Asthma in younger adults • Lower-extremity amputation among patients with diabetes	Discharges for each of the included chronic conditions will decline after implementation of DSRIP.	HCUP data[b]	Medicaid beneficiaries and uninsured individuals	New Jersey Texas
Research question 2: What is the overall effect of DSRIP on use of primary and preventive care?					
Adult ambulatory care visits	Annual percentage of beneficiaries ages 18 to 64 that had an ambulatory care visit	The percentage of adults with ambulatory care visits will increase after DSRIP implementation for demonstration HSAs relative to comparison HSAs (when relevant).	Medicaid administrative data	Medicaid and CHIP beneficiaries	California New York Texas
Children's and adolescents' access to primary care practitioners[d]	Annual percentage of beneficiaries ages 0 to 17 that had a visit with a primary care practitioner	The percentage of children and adolescents with a visit with a primary care practitioner will increase after DSRIP implementation for demonstration HSAs relative to comparison HSAs (when relevant).	Medicaid administrative data	Medicaid and CHIP beneficiaries	California New York Texas

Table II.1 (*continued*)

	Measure description	Hypothesis	Data sources	Population[a]	States with available data
Research question 3: What is the overall effect of DSRIP on use of behavioral health services?					
Behavioral health visits	Annual percentage of beneficiaries that had a behavioral service visit	The percentage of beneficiaries with an outpatient behavioral health visit will increase after DSRIP implementation for demonstration HSAs relative to comparison HSAs (when relevant).	Medicaid administrative data	Medicaid and CHIP beneficiaries	California New York Texas
Research question 4: What is the overall effect of DSRIP on clinical care processes?					
Hemoglobin A1c (HbA1c) testing[d]	Annual percentage of beneficiaries with diabetes that had an HbA1c test in a year	The percentage of adult beneficiaries with diabetes that had an HbA1c test in a year will increase after DSRIP implementation for demonstration HSAs relative to comparison HSAs (when relevant)	Medicaid administrative data	Medicaid and CHIP beneficiaries	California New York Texas
Follow-up within seven days after an ED visit for an ACSC	Quarterly percentage of ED visits for an ACSC that resulted in a follow-up within seven days	The percentage of ED visits for an ACSC that resulted in a follow-up within seven days will increase after DSRIP implementation for demonstration HSAs relative to comparison HSAs (when relevant).	Medicaid administrative data	Medicaid and CHIP beneficiaries	California New York Texas

[a] See Section II.C., Study population, for additional details.

[b] Avoidable ED visits are defined using the New York University Emergency Department visit severity algorithm (bloom et al. 2000a, 2000b).

[c] The numerator of the hospital discharges for ACSCs measure was constructed using HCUP data. The denominator counts were based on data from the American Community Survey.

[d] Measure is a part of the 2019 Core Set of Child or Adult Health Care Quality Measures for Medicaid.

ACSC= ambulatory care-sensitive condition; CHIP = Children's Health Insurance Program; DSRIP = delivery system reform incentive payment; ED = emergency department; HbA1c = Hemoglobin A1c; HCUP = Healthcare Cost and Utilization Project; HSA = hospital service area; PQI = prevention quality indicator.

Research question 1. What is the overall effect of DSRIP demonstrations on shifting care away from ED and inpatient settings?

ED visits. We counted the total number ED visits that did not result in an inpatient stay per 1,000 beneficiaries for each hospital service area (HSA) in a given quarter. If the DSRIP demonstrations increase access to primary care services, use of the ED should decrease. ED use is a commonly constructed measure for DSRIP demonstrations.[7]

Avoidable ED visits. Some ED visits are unavoidable, but a prompt visit to a primary care physician or a specialist could prevent others (Bodenheimer et al. 2002). We measured the total number of potentially avoidable ED visits per 1,000 beneficiaries in a given year. Following Billings et al. (2000a, 2000b), we defined avoidable ED visits as ED visits for (1) conditions that did not require immediate care within 12 hours; (2) conditions that required treatment within 12 hours but that could have been diagnosed and treated in a typical primary care setting; and (3) conditions that required emergency care, but the emergency care could potentially have been avoided with the use of timely and effective primary care (such as flare-ups of asthma or diabetes). If the DSRIP demonstrations increase access to primary care services, the rate of avoidable ED visits should decrease.

Hospital discharges for ACSCs. AHRQ specifies PQIs to identify areas for which good outpatient care can potentially prevent the need for hospitalization (AHRQ 2018). We measured hospital discharges per 100,000 Medicaid and uninsured individuals for the following conditions: diabetes with short-term complications, diabetes with long-term complications, uncontrolled diabetes without complications, diabetes with lower-extremity amputation, chronic obstructive pulmonary disease (COPD), asthma, hypertension, heart failure, or angina without a cardiac procedure. We also examined hospital discharges separately for each condition category (asthma, diabetes, and cardiovascular conditions). This measure reflects the DSRIP demonstrations' focus on transforming care and reducing avoidable hospital use.

Research question 2. What is the overall effect of DSRIP demonstrations on use of primary care and preventive services?

Adult ambulatory care visits. Receiving appropriate ambulatory care, or care provided by health care professionals in outpatient settings, can reduce unnecessary inpatient and ED use. We measured whether adults ages 18 and older had an ambulatory care visit during the year.

Children's and adolescents' access to primary care practitioners. Access to primary care is critical for the health of children and adolescents, and research suggests primary care reduces avoidable ED visits (Bloom 2011). We measured whether children and adolescents ages birth to 17 years had a visit with a primary care practitioner within a year.

---

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

Research question 3. What is the overall effect of DSRIP demonstrations on use of behavioral health services?

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

Research question 4. What is the overall effect of DSRIP demonstrations on clinical care processes?

Comprehensive diabetes care: HbA1c testing. Diabetes is a condition that is highly prevalent among Medicaid beneficiaries, and DSRIP providers commonly select projects that focus on improving care for beneficiaries with diabetes. We measured HbA1c testing among those with diabetes to assess whether DSRIP demonstrations influence the delivery of diabetes care.

Follow-up after discharge from the ED for ACSCs. Standards for high quality care indicate that many patients who visit the ED for ACSCs should have a primary care visit soon afterward.[8] More generally, people who do not receive follow-up care are more likely to be readmitted to the ED (Cook et al. 2004). We measured the share of ED visits for asthma, COPD, hypertension, and diabetes that resulted in a follow-up within seven days of discharge.

B. Data sources

The evaluation used two main data sources: (1) Medicaid administrative enrollment and claims data and (2) hospital discharge data from the HCUP state inpatient databases. (Appendix A, Table A.1 provides a full description of data availability.)

Medicaid administrative data. To examine the impact of DSRIP demonstrations on Medicaid beneficiaries, we used data derived from Medicaid enrollment files and claims paid to providers. We combined three Medicaid administrative data sources because all states transitioned from the Medicaid Statistical Information System (MSIS) to a new, more complex reporting format—the Transformed MSIS, or T-MSIS—during our study period. For periods before a state's transition, we used Medicaid Analytic eXtract (MAX) files, or the early version of MAX known as Alpha-MAX. For periods after a state's transition, we used T-MSIS Analytic Files (TAFs). MAX and Alpha-MAX are both research versions of state MSIS submissions; TAF is a research version of state T-MSIS submissions.[9] These data were available from 2009 through 2017 for California, New York, and Texas. Together, these sources provide uniform and comprehensive data across states for Medicaid and the Children's Health Insurance Program person-level enrollment and service-level claims data.

---

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

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

TAF data are new, and there are several outstanding data quality and reliability questions (see Appendix Table A.2 for results of a TAF data quality assessment). For example, missing geographic information for some states affected the analyses we could conduct and the comparison states we could include. Data limitations in each state influenced our selection of outcome measures and measure construction. For example, California and Texas did not have usable inpatient encounter records for adult beneficiaries at some points during our study period. As a result, we selected only measures that rely on outpatient data (including ED use), which limits our ability to understand whether there was a shift in inappropriate inpatient use at the individual level.[10]

HCUP state inpatient data. We also relied on HCUP data, the largest collection of longitudinal hospital care data in the United States. In contrast to Medicaid administrative data, HCUP data include all people who receive care, including those without insurance—a key target population for DSRIP demonstrations. We used HCUP data for New Jersey and Texas; HCUP data in California and New York were not available in the post-demonstration period.[11] Because these data include only inpatient discharge records from hospitals, we cannot use them to study health outcomes for individuals.

American Community Survey. The U.S. Census Bureau uses the annual American Community Survey (ACS) to collect social, economic, housing, and demographic indicators on the nation, states, counties, and local areas. We used the ACS five-year estimates at the zip code level[12] to construct the denominator of the PQI chronic composite measure (counts of Medicaid beneficiaries and uninsured living within each HSA) and various HSA-level covariates (for instance, mean income).

Dartmouth Atlas of Health Care data. The Dartmouth Atlas of Health Care provides data on hospital and physician capacity, as well as health care use at the national, regional, and local levels based on Medicare data. We used these data to develop HSA-level measures of health care resources, which we then used to identify an appropriate comparison group in relevant models and as a control variable in our regression models using Medicaid administrative data.

Health professional shortage areas. The Health Resources and Service Administration (HSRA) creates health professional shortage area (HPSA) designations to indicate health care provider shortages in primary, mental health, or dental health care. HRSA works with state partners to determine HPSA designations. We used the HPSA data to identify comparison groups and as a control variable in our regression models using Medicaid administrative data.

---

[10] Analyses using HCUP data will enable us to understand aggregate shifts in potentially preventable inpatient use.

[11] At the time of the analysis, HCUP data were only available through 2011 in California and 2015 in New York. Therefore, we were unable to conduct analyses using HCUP data in these states.

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

C. Study population

The DSRIP demonstrations aim to affect care for the entire community across a spectrum of providers. To reflect this, in California and New Jersey, we defined the population eligible for the demonstration as individuals who reside within a catchment area of participating providers, and we used the Dartmouth Atlas hospital service areas (HSAs) to define the hospital catchment areas (Dartmouth Institute for Health Policy and Clinical Practice 2017). In New York and Texas, where the DSRIP demonstrations were implemented across the entire state, we defined the demonstration group as all individuals in the state. Our analyses of Medicaid administrative data included all Medicaid and CHIP beneficiaries up to age 64 who were not dually eligible for Medicare and did not have a disability. Our analyses of HCUP data included all Medicaid beneficiaries and uninsured individuals.

D. Estimating demonstration effects

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 demonstration had not been implemented (the counterfactual). We selected the analytic design and constructed comparison groups separately for each state. The preferred analytic design was a difference-in-differences approach. When identifying a suitable comparison was not feasible for a particular state, we used a simple interrupted time series (ITS) with no comparison group in which we examined changes in the trend of patient-level outcomes before and after the demonstration was implemented.

Medicaid administrative data

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. We did this for two time periods: (1) before and after DSRIP implementation and (2) before and after PRIME implementation.[13]

We matched each demonstration HSA, defined based on having at least one hospital participating in DSRIP within the HSA, to one or more comparison HSAs in California, Virginia, and Washington. We used propensity score matching (Rosenbaum and Rubin 1983) to identify comparison HSAs that were similar to demonstration HSAs based on observed measures of sociodemographic characteristics, health care access, and clinical care processes in the pre-demonstration period. (Appendix A.1 provides a full description of the propensity score matching methods and measures.)

---

[13] Because California limited eligibility to DPHs and was not implemented statewide, we included HSAs within the state that did not have a hospital participating in DSRIP in the comparison group. We drew our comparison group from a pool of states: Connecticut, Delaware, Ohio, Oregon, Utah, Virginia, Washington, and Washington, DC. We included these states because they had sufficient data quality over the course of the study period. Ultimately, only Virginia and Washington had HSAs that were similar to the demonstration HSAs and ended up being included in our regression models.

In New York and Texas, we compared outcomes of interest before and after implementation of DSRIP. Both states implemented their DSRIP demonstrations statewide or in most of the state and did not have a large pool of in-state HSAs that we could use to identify a similar comparison group. Furthermore, potential out-of-state comparison groups with data available for our analysis differed from the intervention HSAs on key characteristics. As a result, we relied on a simple ITS design.

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 regression models to estimate the effect of DSRIP on the outcomes of interest after controlling for HSA or individual characteristics.[14] See Appendix A, Sections A.1 and A.2 for more information on the analytic methods, including descriptions of comparison groups and regression models used to estimate demonstration effects.

HCUP data

In Texas and New Jersey, we compared the outcome of interest before and after implementation of DSRIP. We employed two main analysis methods—we first assessed the trend in hospital discharges for ACSCs over time using unadjusted data. We then used a simple ITS approach to estimate whether the trends in hospital discharges for ACSCs in the pre-demonstration period differed significantly from the demonstration period.

---

[14] In California and New York, we controlled for the following HSA characteristics (measured in 2011, unless otherwise noted): percentage of Medicaid beneficiaries ages 0 through 17; percentage of the population that were white; percentage of the population that were covered by Medicaid, the change in the percentage of the population that were covered by Medicaid from 2011 to 2015; the number of hospital beds per capita; the percentage of Medicaid beneficiaries with a behavioral health condition, the percentage of beneficiaries with one or more chronic conditions (other than a behavioral health condition); the percentage of the population that lived in a primary care shortage area, and the percentage of the population that lived in a mental health shortage area. In Texas, we controlled for the following individual characteristics: sex, age, presence of a behavioral health condition, and presence of one or more chronic conditions (other than a behavioral health condition).

III. RESULTS

This chapter presents multivariate regression results for the outcomes of interest, separately for each state. Appendix B (Figures B.1 through B.25) shows the unadjusted trends in the outcomes of interest, by state.

A. California

In California, we relied on Medicaid administrative data to estimate HSA-level difference-in-differences models for the following outcomes: (1) ED visits, (2) avoidable ED visits, (3) adult ambulatory care visits, (4) primary care visits for children and adolescents, (5) behavioral health visits, (6) HbA1c testing, and (7) follow-up within seven days after an ED visit for ACSCs. Table III.1 summarizes our results.

1.   Shifting care away from the ED

After controlling for community characteristics and seasonality, DSRIP and PRIME were associated with an unfavorable outcome related to ED visits. We found that ED visits in the comparison group decreased after DSRIP implementation, whereas they increased in the demonstration group. This resulted in a difference-in-differences estimate of 8.6 more ED visits per 1,000 beneficiaries than would have otherwise occurred ($p < 0.001$). After PRIME implementation, ED visits increased for both the demonstration and comparison groups, but they increased more in the demonstration group. This resulted in a difference-in-differences estimate of 12.1 more ED visits per 1,000 beneficiaries than would have otherwise occurred ($p < 0.001$).

Similarly, after controlling for community characteristics and seasonality, we found that DSRIP and PRIME were associated with an unfavorable outcome related to avoidable ED visits. Avoidable ED visits decreased after DSRIP implementation for both the demonstration and comparison groups, but they decreased more in the comparison group. This resulted in a difference-in-differences estimate of 5.6 ED visits per 1,000 beneficiaries, an unfavorable outcome ($p < 0.001$). Avoidable ED visits increased after implementing PRIME for both the demonstration and comparison group, but they increased more in the demonstration group. This resulted in a difference-in-differences estimate of 8.0 avoidable ED visits per 1,000 beneficiaries ($p < 0.001$).

2.   Use of primary care and preventive services

After controlling for community characteristics, we found that neither DSRIP nor PRIME were associated with changes in use of primary care or preventive services. These findings were consistent across age groups: for adults and for children and adolescents.

3.   Behavioral health service use

PRIME had a favorable outcome related to behavioral health services. After controlling for community characteristics, we found that the use of behavioral health services increased post-PRIME for both the demonstration and comparison group, but it increased more for the

demonstration group. This resulted in a difference-in-differences estimate of 0.7 ($p < 0.01$), a favorable outcome. The DSRIP results were not statistically significant.

4. Clinical care processes

After controlling for community-level characteristics, we found that neither DSRIP nor PRIME were associated with changes in HbA1c testing among beneficiaries with diabetes. There were no differences in HbA1c testing post-DSRIP for the demonstration or comparison group, whereas post-PRIME, HbA1c testing increased for both groups.

After controlling for community-level characteristics and seasonality, we found that the share of ED visits for ACSCs that resulted in a follow-up visit within seven days remained the same for the comparison group post-DSRIP but decreased in the demonstration group. This resulted in a difference-in-differences estimate of -1.6 percentage points in the probability of having a follow-up visit ($p < 0.05$). However, PRIME was not associated with a change in the share of beneficiaries with an ED visit for ACSCs who had a follow-up visit within seven days.

Table III.1. Impact of DSRIP and PRIME in California from 2009 and 2017 (Medicaid administrative data)

Outcome measure	Number of HSAs	Mean in the pre-period	Demonstration (versus comparison) (95% confidence interval)	DSRIP post-period estimate (95% confidence interval)	DSRIP post*demonstration estimate (95% confidence interval)	PRIME post-period estimate (95% confidence interval)	PRIME post*demonstration estimate (95% confidence interval)
ED visits[a]	61	70.1[b]	-10.6** (-18.2, -3.0)	-6.1*** (-8.9, -3.3)	**8.6*** (4.7, 12.5)**	22.0*** (19.2, 24.8)	**12.1*** (8.2, 16.1)**
Avoidable ED visits[a]	61	40.7[b]	-6.5* (-11.7, -1.4)	-8.0*** (5.2, 10.7)	**5.6*** (2.9, 8.3)**	3.8*** (1.9, 5.7)	**5.6*** (5.2, 10.7)**
Adult ambulatory care visits	61	45.7[c]	0.3 (-2.2, 2.8)	-2.1*** (-3.2, -1.0)	**-0.3 (-1.8, 1.2)**	-2.4*** (-3.5, -1.3)	**-0.01 (-1.6, 1.5)**
Children's and adolescents' access to primary care practitioners	61	60.4[c]	-0.4 (-3.6, 2.8)	0.8 (-1.0, 2.6)	**-0.2 (-2.7, 2.4)**	1.5 (-0.3, 3.3)	**1.0 (-1.6, 3.6)**
Behavioral health visits	61	2.4[c]	-0.3 (-0.7, 1.0)	1.2*** (0.8, 1.5)	**-0.1 (-0.5, 0.4)**	1.2*** (0.9, 1.6)	**0.7** (0.2, 1.1)**
Hemoglobin A1c testing	61	58.3[c]	0.6 (-4.1, 5.3)	2.0 (-0.9, 4.9)	**-0.4 (-4.5, 3.7)**	11.5*** (8.7, 14.4)	**-0.8 (-4.9, 3.3)**
Follow-up within seven days after an ED visit for ACSC	61	17.8[c]	0.7 (-1.3, 2.6)	0.2 (-0.7, 1.1)	**-1.6* (-2.9, -0.3)**	-0.2 (-1.2, 0.7)	**0.5 (-0.8, 1.8)**

Source: Mathematica's analysis of Medicaid administrative data from California, Virginia, and Washington, 2009 to 2017.

Note: Each column represents results from a separate difference-in-differences analysis using linear mixed-effects models. DSRIP post-period is a binary indicator that equals one in the quarters or years after DSRIP was implemented. PRIME post-period is an indicator equal to one in the quarters or years after PRIME was implemented. Notably, because the data were only available through the second year of PRIME, these estimates should be considered preliminary. We constructed the models so that the PRIME post-period estimate is the difference in the outcome of interest relative to the DSRIP post-period (as opposed to the pre-period). Demonstration is a binary indicator variable that equals one for beneficiaries living in California HSAs subject to DSRIP and PRIME; otherwise, it equals zero. The DSRIP post*demonstration interaction term is the main difference-in-differences effect for the DSRIP demonstration, or the difference between the comparison and demonstration groups after DSRIP implementation for the outcome of interest (in bold). The PRIME post*demonstration interaction term is the main difference-in-differences effect for the first two years of PRIME, or the difference between the comparison and demonstration groups after PRIME implementation for the outcome of interest (in bold).

The models also controlled for the following characteristics measured in 2011, the latest year still in the pre-demonstration period for all states: the percentage of beneficiaries who were children, the percentage of beneficiaries with a behavioral health condition, the percentage of beneficiaries with one or more chronic conditions, the percentage of individuals in the HSA who were white, the percentage of individuals in the HSA covered by Medicaid, the percentage difference of individuals in the HSA covered by Medicaid between 2011 and 2015, the percentage difference of individuals who were uninsured in the HSA between the 2011 and 2015, the number of hospital beds per resident in the HSA, the percentage of individuals in the HSA who lived in a primary care shortage area, and the percentage of individuals in the HSA who lived in a mental health shortage area. ED visits, avoidable ED visits, and follow-up after an ED visit for ACSCs are measured at the HSA-quarter level. For quarterly measures, the models controlled for seasonal effects.

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

Table III.1 (*continued*)

[b] The mean is the number of visits per 1,000 beneficiaries.

[c] The mean is a percentage.

*Statistical significance is less than 0.05 but greater than 0.01.

**Statistical significance is less than 0.01 but greater than 0.001.

***Statistical significance is less than 0.001.

ACSC = ambulatory care sensitive condition; DSRIP = delivery system reform incentive payment; ED = emergency department; HSA = hospital service area; PRIME = Public Hospital Redesign and Incentives in Medi-Cal.

B. New Jersey

In New Jersey, we relied on HCUP data to estimate HSA-level ITS models for hospital discharges for ACSCs among Medicaid beneficiaries and uninsured individuals—the target population of the demonstration. We first examined hospital discharges for ACSCs overall and by condition category. Table III.2 summarizes our results.

1.   Shifting care away from inpatient settings

After controlling for seasonality, we found that before DSRIP implementation, the rate of hospital discharges for ACSCs for the overall Medicaid and uninsured population increased (increase of 2.4 visits per 100,000 individuals per quarter, $p < 0.001$). After DSRIP implementation, discharges immediately increased (increase of 109.2 discharges per 100,000 individuals, $p < 0.01$), followed by a rate decrease (a relative decrease of 5.8 visits per quarter, $p < 0.01$). These results varied by condition category:

- Before DSRIP implementation, asthma-related discharges increased (increase of 0.8 visits per 100,000 individuals per quarter, $p < 0.01$). After DSRIP implementation, discharges immediately increased (increase of 42.2 discharges per 100,000 individuals, $p < 0.05$), followed by a rate decrease (a relative decrease of 2.1 visits per 100,000 individuals per quarter compared with the pre-period trend, $p < 0.05$).

- Before DSRIP implementation, diabetes-related discharges increased (increase of 1.0 visit per 100,000 individuals per quarter, $p < 0.001$). After DSRIP implementation, discharges immediately increased (increase of 49.1 visits per 100,000 individuals, $p < 0.01$), followed by a rate decrease (a relative decrease of 2.6 visits per 100,000 individuals per quarter compared with the pre-period trend, $p < 0.05$).

- Before DSRIP implementation, cardiovascular-related conditions increased (0.5 visits per 100,000 individuals per quarter, $p < 0.01$); this rate did not change after DSRIP implementation.

Table III.2. Impact of DSRIP in New Jersey from 2009 to 2015 (HCUP data)

Outcome measure	Number of HSAs	Mean in the pre-period[a]	Time estimate (95% confidence interval)	Post-period estimate (95% confidence interval)	Time*post estimate (95% confidence interval)
Hospital discharges for ACSCs (PQI 92)[b]	41	166.9	2.4*** (1.4, 3.4)	**109.2** (32.4, 185.9)**	**-5.8** (-9.4, -2.3)**
Hospital discharges for specific conditions					
Asthma-related conditions[b]	41	64.4	0.8** (0.3, 1.3)	**42.2* (1.9, 82.4)**	**-2.1* (-4.0, -0.3)**
Diabetes-related conditions[b]	41	49.0	1.0*** (0.7, 1.4)	**49.1** (17.6, 80.5)**	**-2.6*** (-4.0, -1.2)**
Cardiovascular-related conditions[b]	41	53.5	0.5** (0.2, 0.8)	**13.2 (-15.0, 41.3)**	**-0.8 (-2.1, 0.4)**

Source: Mathematica's analysis of HCUP data from New Jersey, 2009 to 2015.

Note: This table presents results from an interrupted time series analysis using linear mixed-effects models. The models include quarterly fixed effects to control for seasonal variation and a random effect for HSAs to control for clustering of repeated observations within HSAs. Time is measured as the number of quarters since DSRIP implementation. Post-period is an indicator variable that equals one in the years after DSRIP was implemented and zero otherwise. The post-period estimate measures whether there was an immediate change in the outcome of interest at the time of DSRIP implementation. The time*post estimate measures whether the trend (or slope) of the outcome trajectory changed after DSRIP was implemented. We fit a separate regression model for each outcome. The columns in bold are the main estimates of interest.

[a] The mean is the number of hospital discharges for ACSCs per 100,000 Medicaid beneficiaries and uninsured individuals.

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

*Statistical significance is less than 0.05 but greater than 0.01.

**Statistical significance is less than 0.01 but greater than 0.001.

***Statistical significance is less than 0.001.

ACSC = ambulatory care sensitive condition; HSA = hospital service area; PQI = prevention quality indicator.

C. New York

In New York, we relied on Medicaid administrative data to estimate HSA-level ITS models for the following outcomes: (1) ED visits, (2) avoidable ED visits, (3) adult ambulatory care visits, (4) primary care visits for children and adolescents, (5) behavioral health visits, (6) HbA1c testing, and (7) follow-up within seven days after an ED visit for ACSCs. Table III.3 summarizes our results.

1. Shifting care away from the ED

After controlling for community characteristics and seasonality, we found that DSRIP was associated with favorable outcomes related to ED visits. ED visits decreased slightly before DSRIP implementation (a decrease of 0.3 visits per 1,000 beneficiaries per quarter, $p < 0.001$). After DSRIP implementation, the number of ED visits immediately fell (a decrease of 28.5 visits per 1,000 Medicaid beneficiaries, $p < 0.001$), followed by a slightly steeper rate decrease (an additional decrease of 0.7 visits per 1,000 beneficiaries per quarter, $p < 0.001$).

We found a similar pattern for avoidable ED visits. After controlling for community characteristics and seasonality, we found that avoidable ED visits were trending down slightly before DSRIP implementation (a decrease of 0.4 visits per 1,000 beneficiaries per quarter, $p < 0.001$). After DSRIP implementation, the number of avoidable ED visits dropped immediately (a decrease of 15.3 visits per 1,000 Medicaid beneficiaries), followed by a slightly steeper rate decrease (an additional decrease of 0.7 visits per 1,000 beneficiaries per quarter, $p < 0.001$).

2. Use of primary care and preventive services

After controlling for community characteristics, we found that DSRIP was associated with a favorable outcome related to adult ambulatory care visits. These visits increased slightly before DSRIP implementation (an increase of 1.3 percentage points per year, $p < 0.001$). After DSRIP implementation, the rate increased slowly but steadily over time (an additional 1.5 percentage-point increase per year relative to the pre-period, $p < 0.001$).

After controlling for community characteristics, we found that the rate of children and adolescents with a primary care visits increased steadily over the study period (an increase of 0.8 percentage points per year, $p < 0.001$). After DSRIP implementation, visits immediately decreased by 15.9 percentage points ($p < 0.001$), but the rate stayed the same.

3. Behavioral health service use

After controlling for community characteristics, we found that DSRIP was associated with an unfavorable outcome related to use of behavioral health services. The percentage of beneficiaries with a behavioral health service visit increased slightly before DSRIP implementation (an increase of 1.1 percentage points per year, $p < 0.001$), but the growth of the trend slowed after DSRIP implementation (small decrease of 0.4 percentage points per year relative to the pre-period, $p < .001$).

Table III.3. Impact of DSRIP in New York from 2009 and 2017 (Medicaid administrative data)

Outcome measure	Number of HSAs	Mean in the pre-period	Time estimate (95% confidence interval)	Post-period estimate (95% confidence interval)	Time*post estimate (95% confidence interval)
ED visits[a]	133	128.0[b]	-0.3*** (-0.3, -0.2)	**-28.5*** (-30.5, -26.5)**	**-0.7*** (-0.9, -0.4)**
Avoidable ED visits[a]	133	65.0[b]	-0.4*** (-0.4, -0.3)	**-15.3*** (-16.4, -14.1)**	**-0.7*** (-0.8, -0.5)**
Adult ambulatory care visits	133	58.3[c]	1.3*** (1.2, 1.4)	**-8.8*** (-9.4, -8.2)**	**1.5*** (1.1, 1.9)**
Children's and adolescents' access to primary care practitioners	133	74.3[c]	0.8*** (0.6, 1.0)	**-15.9*** (-17.0, -14.7)**	**-0.6 (-1.3, 0.0)**
Behavioral health visits	133	7.7[c]	1.1*** (1.1, 1.2)	**-1.1*** (-1.4, -0.9)**	**-0.4*** (-0.6, -0.3)**
HbA1c testing	133	58.3[c]	1.7*** (1.4, 1.9)	**-5.1*** (-6.4, -3.8)**	**0.2 (-0.6, 0.9)**
Follow-up within seven days after an ED visit for ACSC	133	24.0[c]	0.2*** (0.2, 0.3)	**0.3 (-0.7, 1.3)**	**0.1 (0.0, 0.2)**

Source: Mathematica's analysis of Medicaid administrative data from New York, 2009 to 2017.

Note: Each column represents results from a separate interrupted time series analysis using linear mixed-effects models. Time is measured quarterly for ED visits, avoidable ED visits, and follow-up within seven days after an ED visit for ACSCs. Time is measured annually for adult ambulatory care visits, children's and adolescents' access to primary care practitioners, behavioral health service use, and HbA1c testing. Post-period is a binary indicator equal to one in the years after DSRIP was implemented. It measures whether there was an immediate change in the outcome of interest after DSRIP was implemented. The time*post interaction term measures whether the slope of the outcome changed after DSRIP was implemented. The columns in bold are the main estimates of interest.

The models also controlled for the following characteristics measured in 2011: the percentage of beneficiaries who were children, the percentage of beneficiaries with a behavioral health condition, the percentage of beneficiaries with one or more chronic conditions, the percentage of individuals in the HSA who were white, the percentage of individuals in the HSA covered by Medicaid, the percentage difference of individuals in the HSA covered by Medicaid between 2011 and 2015, the percentage difference of individuals who were uninsured in the HSA between 2011 and 2015, the number of hospital beds per resident in the HSA, the percentage of individuals in the HSA who lived in a primary care shortage area, and the percentage of individuals in the HSA who lived in a mental health shortage area. For quarterly measures, the models controlled for seasonal effects.

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

[b] The mean is the number of visits per 1,000 beneficiaries.

[c] The mean is a percentage.

*Statistical significance is less than 0.05 but greater than 0.01.

**Statistical significance is less than 0.01 but greater than 0.001.

***Statistical significance is less than 0.001.

ACSC = ambulatory care sensitive condition; ED = emergency department; HbA1c = hemoglobin A1c; HSA = hospital service area.

4. Clinical care processes

After controlling for community characteristics, we found that the trend in HbA1c testing increased slightly over the study period (an increase of 1.7 percentage points per year, $p < 0.001$). However, the trend did not change after DSRIP implementation.

Likewise, after controlling for community characteristics and seasonality, we found that the trend in follow-up visits within seven days of an ED visit for ACSCs increased slightly over the study period (an increase of 0.2 percentage points per quarter, $p < .001$) and did not change after DSRIP implementation.

D. Texas

In Texas, we relied on Medicaid administrative data and HCUP data to estimate ITS models for the following outcomes: (1) ED visits, (2) avoidable ED visits, (3) hospital discharges for ACSCs, (4) adult ambulatory care visits, (5) primary care visits for children and adolescents, (6) behavioral health visits, (7) HbA1c testing, and (8) follow-up within seven days after an ED visit for ACSCs. We estimated individual-level models when using Medicaid administrative data[15], and HSA-level models when using HCUP data. Tables III.4 and III.5 summarize our results.

1. Shifting care away from ED and inpatient settings

After controlling for individual characteristics and seasonality, we found that the rate of ED visits per 1,000 beneficiaries increased slightly before DSRIP implementation (an increase of 0.9 visits per 1,000 beneficiaries per quarter, $p < 0.001$). After DSRIP implementation, the number of ED visits fell slightly (a decrease of 1.5 visits per 1,000 Medicaid beneficiaries, $p < 0.001$), followed by a slightly steeper decrease over time (a decrease of 0.1 visits per 1,000 beneficiaries per quarter, $p < 0.001$).

After controlling for individual characteristics and seasonality, the rate of avoidable ED visits increased slightly before DSRIP implementation (an increase of 0.1 visits per 1,000 beneficiaries per quarter, $p < 0.001$). After DSRIP implementation, the rate immediately increased (an increase of 5.0 visits per 1,000 beneficiaries, $p < 0.001$), followed by a slight decrease over time (a decrease of 0.7 visits per 1,000 beneficiaries per quarter relative to the pre-period trend, $p < 0.001$).

After controlling for seasonality, we found that hospital discharges for ACSCs for the overall population were flat before DSRIP implementation. After DSRIP implementation, discharges immediately decreased (a decrease of 34.4 visits per 100,000 individuals, $p < 0.01$), followed by a slight rate increase (an increase of 5.2 visits per 100,000 individuals per quarter relative to the pre-period trends, $p < 0.01$). All condition-specific discharge rates followed this general pattern.

---

[15] In the Texas TAF data, a high proportion of enrollment records are missing geographic information (As a result, we were unable to estimate HSA-level models.

2.   Use of primary care and preventive services

After controlling for individual characteristics, we found that the probability of having an ambulatory care visit among adults decreased slightly before DSRIP implementation (a decrease of 0.8 percentage points per year, $p < 0.001$). After DSRIP implementation, the probability immediately decreased by 6.8 percentage points ($p < 0.001$), followed by a slight increase over time (an increase of 1.1 percentage points per year relative to the pre-period, $p < 0.001$).

After controlling for individual characteristics, we found that the probability of having a primary care visit among children and adolescents increased slightly before DSRIP implementation (an increase of 1.1 percentage points per year, $p < 0.001$). After DSRIP implementation, the probability immediately decreased by 4.0 percentage points, followed by a slight decrease over time (a decrease of 1.6 percentage points per year relative to the pre-period, $p < 0.001$).

3.   Behavioral health service use

After controlling for individual characteristics, we found that the probability of having a behavioral health visit increased slightly before DSRIP implementation (an increase of less than 0.1 percentage points per year, $p < 0.001$). After DSRIP implementation, the probability immediately increased by 0.3 percentage points ($p < 0.001$), followed by a slight increase over time (an additional increase of 0.3 percentage points per year relative to the pre-period, $p < 0.001$).

4.   Clinical care processes

After controlling for individual characteristics, we found that the probability of having an HbA1c test among beneficiaries with diabetes increased slightly before DSRIP implementation (an increase of 0.7 percentage points per year, $p < 0.001$). After DSRIP implementation, the probability immediately decreased by 1.6 percentage points ($p < 0.01$), followed by a decrease over time (a decrease of 4.6 percentage points per year relative to the pre-period trend, $p < 0.001$).

After controlling for individual characteristics and seasonality, we found that the probability of having a follow-up within seven days of an ED visit for ACSCs decreased before DSRIP implementation (a decrease of 0.1 percentage points per quarter, $p < 0.001$). After DSRIP implementation, the probability immediately decreased by 1.8 percentage points ($p < 0.001$), followed by an increase over time (an increase of 0.2 percentage points per quarter relative to the pre-period, $p < 0.001$).

Table III.4. Impact of DSRIP in Texas from 2009 to 2017 (Medicaid administrative data)

Outcome measure	Number of beneficiaries	Mean in the pre-period	Time estimate (95% confidence interval)	Post-period estimate (95% confidence interval)	Time*post estimate (95% confidence interval)
ED visits[a]	147,220,000	110.4[b]	0.9*** (0.9, 0.9)	**-1.5*** (-1.8, -1.2)**	**-0.1*** (-0.5, -0.4)**
Avoidable ED visits[a]	147,220,000	67.8[b]	0.1*** (0.1, 0.1)	**5.0*** (4.8, 5.2)**	**-0.7** (-0.7, -0.7)**
Adult ambulatory care visits	7,663,571	60.6[c]	-0.8*** (-0.9, -0.7)	**-6.8*** (-7.0, -6.6)**	**1.1*** (1.1, 1.2)**
Children's and adolescents' access to primary care practitioners	30,278,490	77.6[c]	1.1*** (1.1, 1.2)	**-4.0*** (-4.0, -3.9)**	**-1.6*** (-1.6, -1.6)**
Behavioral health visits	37,942,061	4.1[c]	<0.1*** (0.0, 0.0)	**0.3*** (0.3, 0.3)**	**0.3*** (0.3, 0.3)**
HbA1c testing	147,542	63.1[c]	0.7** (0.2, 1.3)	**-1.6** (-2.8, -0.4)**	**-4.6*** (-5.1, -4.0)**
Follow-up within seven days after an ED visit for ACSC	370,496	23.2[c]	-0.1* (-0.2, 0.0)	**-1.8*** (-2.4, -1.1)**	**0.2*** (0.1, 0.3)**

Source: Mathematica's analysis of Medicaid administrative data from 2009 to 2017.

Note: Each column represents results from a separate interrupted time series analysis using linear mixed-effects models. Time is measured quarterly for ED visits, avoidable ED visits, and follow-up within seven days after an ED visit for ACSCs. Time is measured annually for adult ambulatory care visits, children's and adolescents' access to primary care practitioners, behavioral health service use, and HbA1c testing. Post-period is a binary indicator equal to one in the years after DSRIP was implemented. It measures whether there was an immediate change in the outcome of interest after DSRIP was implemented. The time*post interaction term measures whether the slope of the outcome changed after DSRIP was implemented. The models also controlled for age, sex, the presence of a behavioral health condition, and the presence of one or more chronic conditions. For quarterly measures, the models controlled for seasonal effects. The columns in bold are the main estimates of interest.

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

[b] The mean is the number of visits per 1,000 beneficiaries.

[c] The mean is a percentage.

*Statistical significance is less than 0.05 but greater than 0.01.

**Statistical significance is less than 0.01 but greater than 0.001.

***Statistical significance is less than 0.001.

ACSC = ambulatory care sensitive condition; ED = emergency department; HbA1c = hemoglobin A1c.

Table III.5. Impact of DSRIP in Texas from 2009 to 2017 (HCUP data)

Outcome measure	Number of HSAs	Mean in the pre-period[a]	Time estimate (95% confidence interval)	Post-period estimate (95% confidence interval)	Time*post estimate (95% confidence interval)
Hospital discharges for all chronic conditions (PQI 92)[b]	206	102.8	0.5 (-0.8, 1.8)	**-34.4** (-44.8, -24.0)	**5.2** (3.8, 6.7)
Hospital discharges for specific conditions					
Asthma-related conditions[b]	206	27.8	-0.2 (-0.9, 0.4)	**-10.6** (-16.2, -5.0)	**1.6** (0.8, 2.3)
Diabetes-related conditions[b]	206	31.6	0.7 (-0.2, 1.5)	**-14.9** (-21.5, -8.2)	**1.9** (1.0, 2.8)
Cardiovascular-related conditions[b]	206	41.0	0.1 (-0.6, 0.8)	**-10.3*** (-16.5, -4.2)	**1.7** (0.9, 2.5)

Source: Mathematica's analysis of HCUP data from 2009 to 2017.

Note: This table presents results from an interrupted time series analysis using linear mixed-effects models. The models include quarterly fixed effects to control for seasonal variation and a random effect for HSAs to control for clustering of repeated observations within HSAs. Time is measured as the number of quarters since DSRIP implementation. Post-period is an indicator variable that equals one in the years after DSRIP was implemented and zero otherwise. The post-period estimate measures whether there was an immediate change in the outcome of interest at the time of DSRIP implementation. The time*post estimate measures whether the trend (or slope) of the outcome trajectory changed after DSRIP was implemented. We fit a separate regression model for each outcome. The columns in bold are the main estimates of interest.

[a] The mean is the number of hospital discharges for ACSCs per 100,000 Medicaid beneficiaries and uninsured individuals.

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

*Statistical significance is less than 0.05 but greater than 0.01.

**Statistical significance is less than 0.01 but greater than 0.001.

***Statistical significance is less than 0.001.

ACSC = ambulatory care sensitive condition; HSA = hospital service area; PQI = prevention quality indicator.

IV. DISCUSSION

Overall, our results suggest that the impact of DSRIP varied across states and research questions. We found some statistically significant findings but no overall pattern (Table IV.1).

Table IV.1. Estimated impact of DSRIP: summary

	California (DSRIP)	California (PRIME)	New Jersey	New York	Texas
	Relative change in post-period—DSRIP group	Relative change in post-period—DSRIP group	Change in trend in post-period	Change in trend in post-period	Change in trend in post-period
ED visits	—	—	n.a.	+	+
Avoidable ED visits	—	—	n.a.	+	+
Hospital discharges for ACSCs	n.a.	n.a.	+	n.a.	—
Ambulatory care visits for adults	NS[a]	NS[a]	n.a.	+	+
Primary care visits for children and adolescents	NS[a]	NS[b]	n.a.	NS[a]	—
Behavioral health service use	NS[a]	+	n.a.	—	+
HbA1c testing	NS[a]	NS[a]	n.a.	NS[b]	—
Follow-up after an ED visit for ACSC	—	NS[b]	n.a.	NS[b]	+

Source: Mathematica's analysis of Medicaid administrative data and HCUP data.
Note: For ED visits, avoidable ED visits, and hospital discharges for ACSCs, an increase in the outcome of interest signifies a negative outcome, whereas a decrease signifies a positive outcome.

[a] Although not statistically significant, results suggest that DSRIP was associated with an unfavorable outcome.
[b] Although not statistically significant, results suggest that DSRIP was associated with a favorable outcome.

+ = favorable outcome; — = unfavorable outcome; NS = not significant; n.a. = not applicable; ACSC = ambulatory care sensitive condition; DSRIP = delivery system reform incentive payment; ED = emergency department; HbA1c = hemoglobin A1c; PRIME = Public Hospital Redesign and Incentives in Medi-Cal.

In California, findings through the first two years of PRIME were largely nonsignificant or negative. Both DSRIP and PRIME had negative effects related to shifting care away from the ED, and neither demonstration was associated with changes in the use of primary or preventive care. PRIME was linked to an increase in the share of beneficiaries with a behavioral health visit, but the magnitude of the difference was small (a less than 1 percentage point relative difference between HSAs). Neither DSRIP nor PRIME was meaningfully associated with improvements in the measures of clinical care processes; after implementing DSRIP, there was a one percentage point relative decrease in the probability of having a follow-up after an ED visit for an ACSC compared to the pre-period, and all other findings were not statistically significant.

In New Jersey, DSRIP was linked to a positive shift in care away from inpatient settings, as measured by hospital discharges for ACSCs.

New York demonstrated positive results related to the first two research questions. DSRIP was associated with a shift in care away from the ED, although the magnitude of the difference was

small (a decrease of less than one visit per 1,000 beneficiaries per quarter after DSRIP implementation for both general and avoidable ED visits). This shift in ED visits also coincided with a rise in the share of adult beneficiaries with an ambulatory care visit, but the magnitude was also small (about a 1 percentage point increase per year after DSRIP implementation). However, DSRIP was negatively associated with the use of behavioral health services (half a percentage point decrease per year), and there were no significant findings related to clinical care processes.

Findings in Texas were mixed. For instance, DSRIP was associated with favorable outcomes related to shifting care away from the ED but unfavorable outcomes related to shifting care away from inpatient settings. Furthermore, DSRIP was associated with an increase in the probability of having an ambulatory care visit for adults, a behavioral health visit, and a follow-up within seven days after an ED visit for an ACSC, but it was associated with a decrease in the probability of having a primary care visit for children and adolescents and having an HbA1c test. Although all findings in Texas were significant (likely because of the number of beneficiaries in the models), the magnitude of the findings was very small.

A. Study limitations

In evaluating DSRIP's effects on access to care and clinical care processes, we faced several challenges related to data, concurrent changes in Medicaid policy, and limitations of the specific methods used. As described below, we addressed these challenges to the extent possible.

Data limitations. The transition of Medicaid data systems from MSIS to T-MSIS, which occurred during the study period for all states, greatly influenced our analyses relying on these data. Texas implemented T-MSIS in July 2014, New York in July 2015, and California and New Jersey in October 2015.

Although these new data should be timelier and of higher quality than prior data, there are still important data quality concerns (see Appendix Table A.2). Most strikingly, we saw a large dip in the number of claims in the quarter just before each state switched to T-MSIS. For our quarterly measures, we dropped all observations in that quarter so that this under-representation of services would not bias the estimates. But we could not address this data anomaly in our annual measures, so our numerators and denominators for the year that a state made the transition to T-MSIS are lower than expected. This is particularly problematic in California and New York, which implemented PRIME and DSRIP, respectively, at about the same time that they switched to T-MSIS. Given these concurrent data system changes, it is difficult to determine whether changes in the outcomes of interest are the result of DSRIP or of the switch to a new data system.

Besides the limitations of the Medicaid administrative data, HCUP data also have important limitations. For example, HCUP data include only a limited set of individual characteristics, such as age and gender. This hinders our ability to control for patient mix, which can have important implications for estimating the impact of DSRIP on the outcomes measured. HCUP data also include only information related to inpatient discharges. As a result, we cannot use these data to

assess the extent to which DSRIP transformed the delivery system as a whole—a key goal of the DSRIP demonstrations.

Finally, on October 1, 2015, U.S. hospitals changed their diagnosis and procedure coding systems, transitioning from the *International Classification of Diseases,* 9th edition, Clinical Modification/Procedure Classification System (ICD-9-CM/PCS) to ICD-10-CM/PCS codes. The ICD-10-CM diagnosis codes provide greater detail in areas such as acuteness and related comorbidities, making one-to-one matching between many diagnosis codes infeasible. To account for this, AHRQ updated the specifications for the PQI-92 composite measure to accommodate changes to diagnostic and procedural coding systems. This specification change became effective January 1, 2015. For this reason, trend changes that begin in 2015 may be due to changes in the measure specifications. The impact of this update is particularly apparent in Texas, where our post-period includes data from 2015 through 2017.

Concurrent changes in Medicaid policy. The demonstrations are unfolding in the context of a rapidly changing health care system, and many forces beyond a demonstration affect the outcomes of interest—circumstances that make isolating the impact of DSRIP a challenge. For instance, California, New Jersey, and New York expanded Medicaid eligibility to low-income adults through the ACA in 2014. Isolating a DSRIP effect is particularly difficult in New Jersey, which implemented DSRIP and expanded Medicaid eligibility simultaneously.

Methods limitations. This summative evaluation included both simple ITS and difference-in-differences analyses, each of which has important limitations. A simple ITS model assumes that, in the absence of the demonstration, trends in the outcomes of interest during the pre-period would have remained the same, which may not be the case. ITS models also assume that changes in trends that occur after DSRIP implementation can be attributed to the demonstration. But this may not be true; concurrent changes in Medicaid policy, as mentioned above, may be driving our results. Difference-in-differences models, although the preferred approach, also have significant limitations. Most important, the validity of these models depends on identifying a similar comparison group. These models also assume that the decision to deploy the demonstration is not related to the outcomes of interest before implementation—for example, that high use of acute care and low use of primary care did not drive a state's decision to implement DSRIP. This is likely not the case.

To the extent possible, we attempted to address these limitations. In California, we attempted to isolate the impact of DSRIP by selecting a comparison group that, to the extent possible, was affected by similar forces as the demonstration group. Our results suggested that including a comparison group was important for identifying the impact of DSRIP and PRIME. In addition to our difference-in-differences models, we conducted sensitivity analyses using interrupted time series models. The results of these two models were inconsistent at times (see Appendix Table C.1). For instance, in the difference-in-differences models, we found insignificant findings related to ambulatory care visits, primary care visits, HbA1c testing, and follow-up within seven days after an ED visit for ACSCs for either DSRIP or PRIME. However, in the ITS models, these findings were significant. This suggests that factors other than the DSRIP demonstration

affected the trajectory of the key outcomes of interest in both the demonstration and comparison groups. As a result, ITS results from New Jersey, New York, and Texas should be interpreted with caution.

Although including a comparison group improved our estimates in California, these analyses still have important limitations, as described above. For instance, DSRIP HSAs differ from non-DSRIP HSAs in some ways that we could not control for in our analysis. In California, the comparison HSAs consisted of communities that did not implement DSRIP or PRIME. Only DPHs and district and municipal public hospitals implemented these programs, whereas comparison HSAs were served by other types of hospitals. Demonstration and comparison hospitals and the communities they serve can differ in several unobserved ways, which could also differentially affect the outcomes observed there. Furthermore, comparison communities outside California operate within different Medicaid programs and are affected by different state policies and regulations. However, by matching demonstration hospitals to comparison hospitals on observable attributes and reviewing baseline trends across both groups, we were able to maximize the validity of the comparisons. In addition, our analysis of PRIME is preliminary as data cover only the first two years of PRIME.

Finally, although DSRIP demonstrations are implemented by providers that serve a large portion of a community or state, often only a small share of Medicaid beneficiaries and uninsured individuals are directly exposed to activities carried out under DSRIP. We are not able to identify those beneficiaries that are directly exposed to the demonstration. As a result, our study may not be able to detect differences in the outcomes of interest that are due to DSRIP.

B. What have other studies found?

The findings from this study fill a critical gap in understanding the impact of DSRIP demonstrations on Medicaid beneficiaries and the uninsured. To date, only one study, California's interim evaluation of the PRIME program, used administrative data to assess the impact of DSRIP demonstrations overall on outcomes of interest over a sufficient period before and after demonstration implementation. Although the interim evaluation of PRIME relied on administrative data, the analysis focused on comparing results from participating hospitals to similar non-participating hospitals. This approach differs from the analysis presented here, which compared results from participating HSAs to comparison HSAs.[16] As a result, our findings cannot be directly compared with findings from other studies. However, the results from state interim and final evaluations provide helpful context about the extent to which DSRIP demonstrations, and more commonly, implementation of specific projects, are influencing milestones and metrics of interest (Table IV.2).

---

[16]Although comparing participating HSAs to comparison HSAs may underestimate the impact of a specific hospital or hospital system, it does have advantages over a hospital-level analysis. By including all beneficiaries in an HSA, not just those served by a participating hospital, this analysis addresses whether DSRIP and PRIME impact care and care processes for the target population more broadly.

Table IV.2. Comparison of DSRIP summative evaluation findings to state final and interim evaluation findings, by research question

Research question	California (DSRIP)		California (PRIME)		New Jersey		New York		Texas	
	Summative evaluation	State final evaluation	Summative evaluation	State interim evaluation	Summative evaluation	State interim evaluation	Summative evaluation	State interim evaluation	Summative evaluation	State final evaluation
What is the overall effect of DSRIP demonstrations on shifting care away from emergency department and inpatient settings?	—	Mixed	—	—	+	Mixed	+	—	Mixed	n.a.
What is the overall effect of DSRIP demonstrations on use of primary and preventive services?	NS	Mixed	NS	Mixed	n.a.	NS	Mixed	NS	Mixed	n.a.
What is the overall effect of DSRIP demonstrations on use of behavioral health services?	NS	n.a.	+	n.a.	n.a.	Mixed	—	+	+	n.a.
What is the overall effect of DSRIP demonstrations on clinical care processes?	Mixed	Mixed	NS	Mixed	n.a.	n.a.	NS	Mixed	Mixed	n.a.

Source: Mathematica's analysis of Medicaid data from 2009 to 2017 and HCUP data from 2009 to 2017.

Pourat, Nadereh, Ying-Ying Meng, Arleen Leibowitz, Jack Needleman, Xiao Chen, Dylan H. Roby, et al. "Final Evaluation Report of California's Delivery System Reform Incentive Payments (DSRIP) Program." 2016. Available at http://healthpolicy.ucla.edu/publications/Documents/PDF/2017/dsrip-report-jun2017.pdf.

Pourat, Nadereh, Xiao Chen, Ana E. Martinez, Lina Tieu, Maria Ditter, Michael P. Huynh, Leigh Ann Haley, Jack Needleman, Denisse M. Huerta. "Interim Evaluation of California's Public Hospital Redesign and Incentives in Medi-Cal (PRIME) Prgram." 2020. Available at http://healthpolicy.ucla.edu/publications/Documents/PDF/2019/PRIME-Interim-Report-Final.pdf.

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.

Table IV.2 (*continued*)

Weller, W., K. Bailey, S. Chatterjee, S. Drazek, R. Greene, M. Gullick, K. Lahiri, E. Martin, Y.J. Park, S. Rain, D. Roby, and A. Rozsavolgyi. "Final Interim Evaluation Report by the Independent Evaluator for the New York State Delivery System Reform Incentive Payment Program." August 2019. Available at https://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/ny/medicaid-redesign-team/ny-medicaid-rdsgn-team-dsrip-final-interim-eval-20190802.pdf

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.

+ = favorable outcome; — = unfavorable outcome; Mixed = results across measures within a research question were mixed; NS = not significant; n.a. = not applicable.

California. California's final evaluation of its initial DSRIP demonstration and the interim evaluation of the PRIME program provide important context for understanding the results of this report. The final DSRIP evaluation drew heavily on descriptive quantitative analyses, qualitative data, and self-reported milestone achievement among participating DPHs (Pourat et al. 2016). The evaluators used a difference-in-differences framework to compare DPHs that implemented specific projects to those that did not, using aggregate DPH data. The evaluators also measured changes in aggregate DPH performance over time and compared these changes to regional and national benchmarks.

Their findings suggest that DSRIP projects had mixed results; some projects achieved improved performance on certain clinical quality measures, but others had declines in performance or no significant change. For instance, the evaluators found that DPHs implementing care transitions projects had significantly better performance on 6 of 13 measures,[17] and those implementing chronic care management projects had significantly better performance on 4 of 13 measures. However, implementing medical home projects was associated with unfavorable changes in several areas: getting timely appointments, uncontrolled diabetes, influenza immunization, tobacco cessation, diabetes: low-density lipoprotein (LDL) control, diabetes: HbA1c control, and optimal diabetes care.

When looking at descriptive trends, the evaluators found modest improvements in all four care coordination measures (diabetes complications, uncontrolled diabetes, congestive heart failure [CHF] admissions, and COPD admissions) and in most (five of six) at-risk patient population measures (diabetes: LDL control, diabetes: HgA1c control, CHF readmissions, hypertension control, pediatric asthma, and an optimal diabetes care composite)[18] during the demonstration period. DPHs also improved on the preventive health measures examined.

Although the findings from our study were mixed, they were overall more favorable than the findings in the state's final evaluation report. This may be because the final evaluation report for California did not identify a comparison group that was unaffected by DSRIP.

California's interim evaluation of PRIME also used a difference-in-differences framework to compare outcomes for patients served by participating DPHs and district municipal public hospitals (DMPHs) with those of patients served by similar comparison hospitals. The evaluators examined 18 measures that assessed the extent to which participating hospitals achieved three demonstration goals: (1) increased provision of patient-centered, data-driven, team-based care; (2) improved provision-of-care services, complex care management, population health management, and culturally competent care; and (3) improved population health and patient experience in Medi-Cal. Evaluators found that outcomes among DPHs tended to improve for primary care, complex care management, and population health management, whereas outcomes

---

[17] The 13 measures consisted of 5 Clinician and Group Consumer Assessment of Healthcare Providers and Systems measures and 8 clinical care processes and outcome measures. DPHs that did not implement care transitions projects showed better performance on short-term diabetes complications and the optimal diabetes care composite measure, whereas the other measures did not show significant changes.

[18] One measure (hypertension control) remained the same.

were generally mixed among DMPHs. By contrast, our findings suggested that primary care and clinical care processes did not change over the first two years of PRIME. The evaluators also found that both DPHs and DMPHs experienced worse outcomes related to population health and patient experience, which was consistent with our findings.

New Jersey. The midpoint assessment in New Jersey (Chakravarty et al. 2014), which examined the impact of DSRIP from 2011 to 2013,[19] relied on an analysis of Medicaid claims data to examine patient care, health, costs, and hospitals' finances. The analysis included data only from 2011 through 2013, before DSRIP implementation. Over the time period examined, some measures improved (such as avoidable hospitalizations for asthma and diabetes); others worsened (such as ED visits for asthma among adults); and many had no changes (such as follow-up after a hospitalization for mental illness and 30-day readmissions for heart failure, acute myocardial infarction, pneumonia, and COPD).

Similar to our findings, this report showed that avoidable hospitalizations for asthma and diabetes improved (the discharges decreased) after implementing DSRIP.

New York. The interim evaluation for New York included a qualitative assessment of program implementation, along with a quantitative analysis of changes in health care quality, population health, behavioral health service use, and avoidable hospital utilization (Weller et al. 2019).

Drawing on data from 2014 to 2017, the quantitative analysis yielded mixed results. The analysis of health care quality focused on two measures of asthma care, one that improved and one that did not change. The analysis of population health outcomes focused on 11 statewide measures. Six of the 11 measures showed improvement, including measures of premature death and people in poor mental health, but the evaluator did not test for statistical significance. Statewide, three of four measures of behavioral health care use either increased immediately after DSRIP implementation or increased over time in the post-period, particularly among large PPSs. Finally, the trend in potentially preventable hospital readmissions and potentially preventable ED use did not change.

Although the measures in the New York interim evaluation did not directly correspond to those used in this report, we did investigate some similar research questions. However, the findings from the two reports were generally inconsistent. In some cases, the state found positive results where we found negative results, and vice versa. For instance, New York found largely positive trends related to behavioral health service use, but we found that these trends decreased after DSRIP implementation. The New York interim evaluation also showed that DSRIP was not associated with reductions in ED visits, whereas we saw decreases in ED visits and avoidable ED visits. These inconsistent findings may be a result of differences in the length of the pre-period (our study focused on pre-period trends starting in 2009, whereas the New York interim evaluation only included data starting in 2014). The inconsistencies may also be due to differences between the state administrative data and the TAF.

---

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

Texas. To assess the impact of Texas's initial DSRIP demonstration on quality of care, health, and costs, evaluators used a comparative case study of 10 care navigation projects (Texas Health and Human Services 2017; Smith 2017).[20] The study included site visits, interviews, survey data, and a limited analysis of hospital discharge data. The evaluators compared findings from providers implementing these projects with findings from 10 providers that did not implement these projects. Overall, the evaluators found that care navigation projects did not improve outcomes relative to comparison providers. Given the scope of this analysis, the findings are not comparable to those in this report.

C. Conclusion

We sought to understand the impact of DSRIP demonstrations on key outcomes in the first four states that implemented these demonstrations. Although the results were mixed, we did see some promising findings—particularly in New York, the most ambitious program described in this report. Although these findings are encouraging, we could not identify a reasonable comparison group for the communities in New York, which prevents us from conclusively attributing our findings directly to DSRIP.

More broadly, this study underscores the challenges of isolating the impacts of broad Medicaid programs aimed at achieving the triple aim of better care, lower costs, and improved health. First, DSRIP demonstrations are only one of many initiatives that states are implementing, and attributing any changes in outcomes to one program is difficult. Further, other health system changes, such as a transition to a new data or procedure coding system, likely affect the results but are not due to DSRIP implementation. Finally, the most rigorous method for estimating causal effects in non-experimental studies requires identifying a reasonable comparison group. This is difficult and sometimes infeasible when considering out-of-state comparison groups because Medicaid program characteristics—for instance, financing, eligibility, and benefit design—vary considerably from state to state.

---

[20] The final evaluation of the Texas Demonstration Waiver broadly focused on Medicaid managed care expansion, uncompensated care, regional safety care structures, and stakeholders' perceptions; DSRIP was only a small component of the report.

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

REFERENCES

AHRQ. "Prevention Quality Chronic Composite: Technical Specifications." Version 2018. Rockville, MD: AHRQ, June 2018. Available at https://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V45/TechSpecs/PQI%2092%20Prevention%20Quality%20Chronic%20Composite.pdf. Accessed December 30, 2019.

Baller, Julia B., James Woerheide, Keanan Lane, Natalya Verbitsky-Savitz, and Marian V. Wrobel. "Delivery System Reform Incentive Payments: Medicaid 1115 Demonstration Interim Evaluation Report." Final report to CMS. Washington, DC: Mathematica Policy Research, January 2018. Available at https://www.medicaid.gov/medicaid/downloads/dsrip-interim-eval-report.pdf. Accessed December 30, 2019.

Baller, Julia, Natalya Verbitsky Savitz, and Jessica Heeringa. "Delivery System Reform Incentive Payments: Final Evaluation Design Plan." 2019. Available at https://www.medicaid.gov/medicaid/downloads/final-eval-dsgn-dlvry-incntv-pyments.pdf. Accessed December 26, 2019.

Billings, John, Nina Parikh, and Tod Mijanovich. "Emergency Department Use: The New York Story." New York: Commonwealth Fund, November 2000a. Available at https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_issue_brief_2000_nov_emergency_room_use__the_new_york_story_billings_nystory_pdf.pdf. Accessed December 26, 2019.

Billings, John, Nina Parikh, and Tod Mijanovich. "Emergency Department Use in New York City: A Substitute for Primary Care?" New York: Commonwealth Fund, November 2000b. Available at https://www.commonwealthfund.org/publications/issue-briefs/2000/mar/emergency-department-use-new-york-city-substitute-primary-care. Accessed December 26, 2019.

Bloom, B., R.A. Cohen, and G. Freeman. "Summary Health Statistics for U.S. Children: National Health Interview Survey, 2011." *Vital Health Statistics*, series 10, no. 254. Hyattsville, MD: National Center for Health Statistics, 2012. Available at http://www.cdc.gov/nchs/data/series/sr_10/sr10_254.pdf. Accessed December 26, 2019.

Bodenheimer, T., E.H. Wagner, and K. Grumbach. "Improving Primary Care for Patients with Chronic Illness: The Chronic Care Model, Part 2." *JAMA*, vol. 288, no. 15, 2002, pp. 1909–1914.

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. Available at December 26, 2019.

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.

Han, B., W.M. Compton, C. Blanco, and L.J. Colpe. "Prevalence, Treatment, and Unmet Treatment Needs of U.S. Adults with Mental Health and Substance Use Disorders." *Health Affairs*, vol. 36, no. 10, 2017, pp. 1739–1747.

Lane, Keanan, Raga Ayyagari, and Jessica Heeringa. "Building Provider Capacity for Delivery System Transformation: The Role of Section 1115 Delivery System Reform Incentive Payment Demonstrations." Final report to CMS. Cambridge, MA: Mathematica, 2019.

Pourat, Nadereh, Ying-Ying Meng, Arleen Leibowitz, Jack Needleman, Xiao Chen, Dylan H. Roby, et al. "Final Evaluation Report of California's Delivery System Reform Incentive Payments (DSRIP) Program." 2016. Available at http://healthpolicy.ucla.edu/publications/Documents/PDF/2017/dsrip-report-jun2017.pdf. Accessed December 26, 2019.

Pourat, Nadereh, Xiao Chen, Ana E. Martinez, Lina Tieu, Maria Ditter, Michael P. Hunyh, et al. "Interim Evaluation of California's Public Hospital Redesign and Incentives in Medi-Cal (PRIME)." 2020. Available at http://healthpolicy.ucla.edu/publications/Documents/PDF/2019/PRIME-Interim-Report-Final.pdf. Accessed August 25, 2020.

Rosenbaum, Paul R., and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, vol. 70, no. 1, 1983, pp. 41–55.

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. Accessed December 26, 2019.

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. Accessed December 26, 2019.

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. Accessed December 26, 2019.

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 & Medicaid Services and California Health and Human Services Agency. "Special Terms and Conditions, No. 11-W-00193/9, California Medi-Cal 2020 Demonstration." Approved December 31, 2015 through December 31, 2020.

U.S. Centers for Medicare & 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 & 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 New York State Department of Health. "Special Terms and Conditions, No. 11-W-00114/2, New York Partnership Plan Section 1115 Medicaid Demonstration." Approval Period: August 1, 2011 – December 31, 2014; as amended April 14, 2014.

U.S. 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." Approval period: December 12, 2011, through September 30, 2016; amendment approved February 26, 2015.

U.S. 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." Approval period: January 2, 2018, through September 30, 2022; approved December 21, 2017.

Walker, E.R., J.R. Cummings, J.M. Hockenberry, and B.G. Druss. "Insurance Status, Use of Mental Health Services, and Unmet Need for Mental Health Care in the United States." *Psychiatric Services*, vol. 66, no. 6, 2015, pp. 578–584.

Weller, W., K. Bailey, S. Chatterjee, S. Drazek, R. Greene, M. Gullick, K. Lahiri, E. Martin, Y.J. Park, S. Rain, D. Roby, and A. Rozsavolgyi. "Final Interim Evaluation Report by the Independent Evaluator for the New York State Delivery System Reform Incentive Payment Program." August 2019. Available at https://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/ny/medicaid-redesign-team/ny-medicaid-rdsgn-team-dsrip-final-interim-eval-20190802.pdf . Accessed December 26, 2019.

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

Appendix A:

Methods

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

SECTION A.1. HOSPITAL SERVICE AREA MATCHING APPROACH AND RESULTS IN CALIFORNIA[21]

In California, we used propensity score matching to identify a set of comparison hospital service areas (HSAs) that were similar to the demonstration HSAs on a range of sociodemographic, health care access, and other key covariates measured before the implementation of the demonstration. Matching on the estimated propensity score, or the estimated probability that an HSA would have participated in the demonstration (if available) given the observed characteristics, enabled us to efficiently balance the demonstration and comparison groups on a large set of covariates and to produce two groups that were similar on the *observed* characteristics. The rest of this section describes our matching approach.

We performed propensity score matching in five steps. First, we identified outliers in the comparison group and dropped them from the analysis. If the outlier value was two or more standard deviations greater than the maximum value in the treatment group, we dropped it. Next, we assessed overlap (or common support) between the demonstration and potential comparison group HSAs on key pre-demonstration covariates and dropped potential comparison HSAs that lacked overlap. Third, we estimated propensity scores for each HSA using a logistic regression model of the following form:

$$(1) \qquad \log\left[\frac{P(Y_j = 1)}{1 - P(Y_j = 1)}\right] = \beta_0 + \beta_1 X_j,$$

where $P(Y_j = 1)$ is the expected probability that an HSA would have participated in the demonstration based on the observed covariates (X_j). We included pre-demonstration trends (slopes) and levels (intercepts) of the outcomes in the propensity score model to account for the pre-intervention trends in the outcomes, which are expected to be the best predictors of how these outcomes will behave in the post-intervention period in the absence of the intervention.

Fourth, we matched each demonstration HSA to up to three comparison HSAs. We matched with replacement, meaning one demonstration HSA could be matched to up to three comparison units and several demonstration HSAs could be matched to the same comparison unit. To account for the possibility of multiple matches, we created weights that we used in the analysis. Finally, we assessed balance of each covariate before and after matching. We repeated these four matching steps separately for physical and behavioral health outcomes. We drew all comparison groups from the same eight states (Connecticut, Delaware, Ohio, Oregon, Utah, Virginia, Washington State, and Washington, DC), which did not participate in the DSRIP demonstration during the study period and had data available to support the impact analysis. Ultimately, we only identified similar HSAs in Virginia and Washington. The resulting samples show balance on most of the key characteristics (Figure A.1).

---

[21] We attempted to find appropriate comparison groups in New York and Texas, but we were unable to find similar communities.

Figure A.1. Pre-and post-matching balance plot of potential HSA-level characteristics and slopes for California models
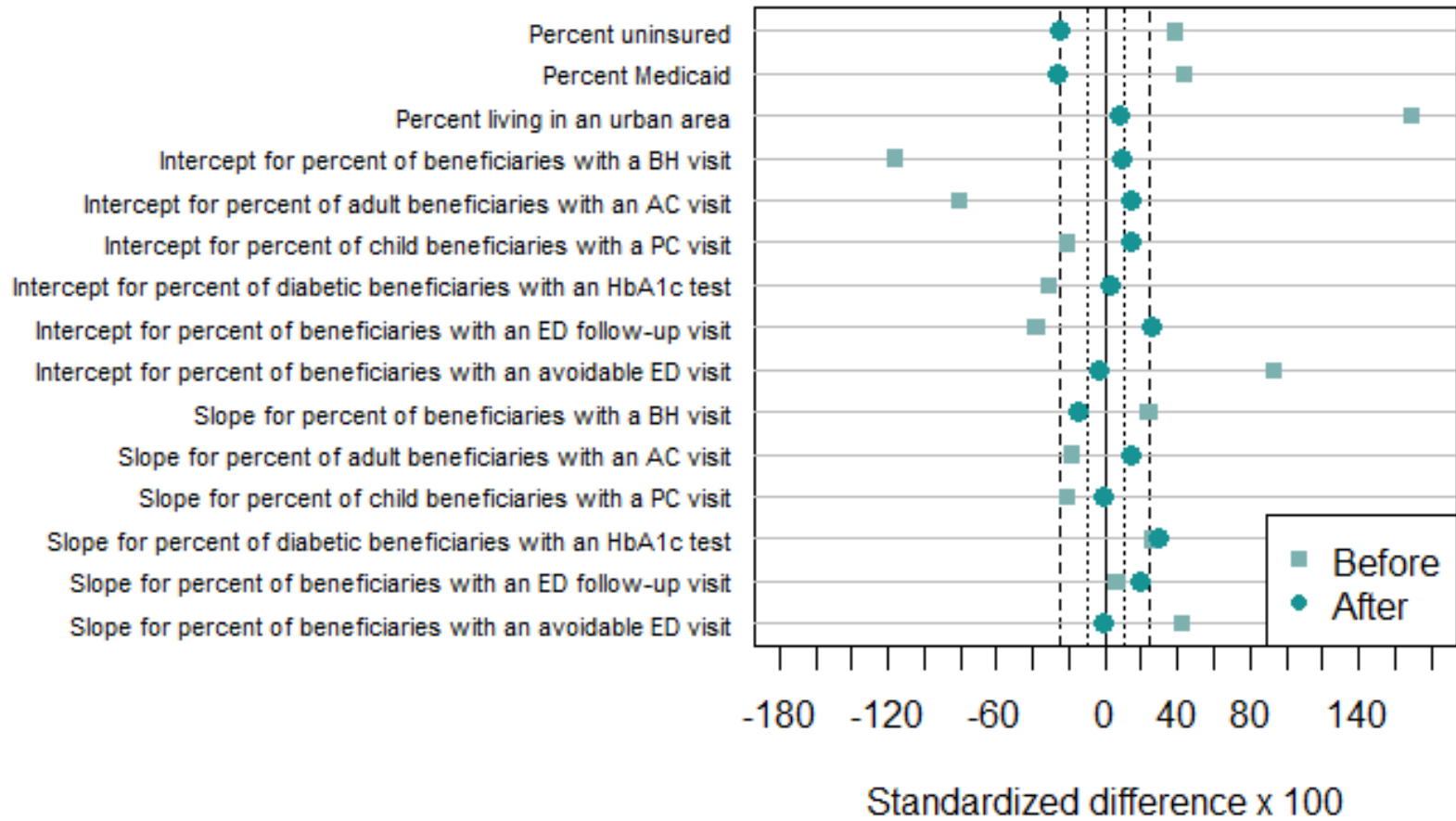
AC = ambulatory care; ACSC = ambulatory care sensitive condition; BH = behavioral health; ED = emergency department; HbA1c = hemoglobin A1c; HSA = hospital service area; PC = primary care.

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

State	Implementation start date	MAX	Alpha-MAX	TAF	HCUP
Demonstration states					
California	November 1, 2011 (DSRIP); July 1, 2016 (PRIME)	2009–2014	January 2015–September 2015	October 2015–December 2017	—
New Jersey	January 1, 2014	2009–2014	January 2015–September 2015	October 2015–December 2017	2009–2015
New York	April 1, 2015	2009–2014	January 2015-June 2015	July 2015 December 2017	—
Texas	October 1, 2011	2009–2013	January 2014–June 2014	July 2014–December 2017	2009–2017
Comparison states					
Virginia	N/A	2009–2014	January 2014–March 2014	January 2014–December 2017	—
Washington	N/A	2009–2014	—	2015-2017	—

DSRIP = delivery system reform incentive payment; HCUP = Healthcare Cost and Utilization Project; MAX = Medicaid Analytic eXtract; N/A = Not Applicable; PRIME = Public Hospital Redesign and Incentives in Medi-Cal; TAF = T-MSIS Analytic Files.

Table A.2. T-MSIS Analytic File level of data quality concerns

	California (2015-2017)	New York (2014-2017)	Texas (2014-2017)	Virginia (2014-2017)	Washington (2015-2017)
Medicaid and CHIP enrollment[a]	Low	Medium	Medium	Low	High[b]
Claims linking to enrollment records—OT file[c]	Low	Low	Low	Low	Low
Claims linking to enrollment records—RX file[c]	Low	Low	Low	Low	Low
Volume of claims—OT file[d]	Medium	Low	Low	Low	Low
Volume of claims—RX file[d]	Low	Low	Low	Low	Low
Encounter data usability—OT file[d]	Medium	Medium	Low	Medium	Low
Encounter data usability—RX file[d]	Medium	Medium	Low	Low	Low
Diagnosis codes—OT file[e]	Low	Low	Low	Low	Low
Procedure codes—RX file[e]	Low	Medium	Low	Low	Low

Source: Mathematica analysis of TAF version 3.0 as of October 2019.

Notes: We looked across all years of data available for each state. We considered a state to have a low data quality concern if it met criteria for all years. We considered a state to have a medium data quality concern if it met criteria for at least one year but had no years in which it met criteria for a high data quality concern. We considered a state to have a high data quality concern if it met criteria for at least one year.

[a] To determine the level of data quality concern for Medicaid and CHIP enrollment, we compared the number of Medicaid and CHIP beneficiaries with comprehensive benefits in the 2016 TAF to the Eligibility and Enrollment Performance indicator (PI) data. We calculated the percent difference between the two. If the TAF counts were within 10 percent of the PI counts, we considered there to be a low data quality concern. If the TAF counts were between 10 to 20 percent of the PI counts, we considered there to be a medium data quality concern.

[b] Although Washington had a high data quality concern regarding benchmarking full-benefit Medicaid and CHIP enrollment, we can use the CHIP code to successfully identify the population of interest. As a result, we opted to include Washington as a comparison state.

[c] To determine the level of data quality concern for claims linking to enrollment records, we calculated the percentage of claims that successfully linked to enrollment records. We considered a state to have a low data quality concern if more than 95 percent of claims successfully linked to enrollment records.

[d] To determine the level of data quality concern for volume of claims and encounter data usability, we calculated the number of total header records or the number of encounter header records per 1,000 enrolled months. We compared the state's volume to the national median. We considered a state to have a low data quality concern if the state fell within 75 to 150 percent of the national median. We considered a state to have a medium data quality concern if the state fell within 150 to 200 percent or 50 to 75 percent of the national median.

[e] To determine the level of data quality concern for diagnosis codes and procedure codes, we examined the percent of records with missing or invalid codes. We considered a state to have a low data quality concern if less than 10 percent of records were missing or invalid. We considered a state to have a medium data quality concern if between 10 and 20 percent of records were missing or invalid.

CHIP = Children's Health Insurance Program; OT = other services; RX = prescription drug; T-MSIS = Transformed Medicaid Statistical information System.

SECTION A.2. ESTIMATING DEMONSTRATION EFFECTS

A. Outcomes Derived from Medicaid administrative data

California: difference-in-differences

In California, we used a difference-in-differences approach to test the causal effects of Delivery System Reform Incentive Payment (DSRIP) and the Public Hospital Redesign and Incentives in Medi-Cal (PRIME) demonstrations on patient outcomes. This econometric technique enabled us to determine whether outcomes have changed differently for the demonstration group than for the comparison group at two points in time: (1) after implementation of DSRIP; and (2) after implementation of PRIME. Specifically, the HSA-level models estimated the differences in the regression-adjusted changes at two time points (from the pre-period to after DSRIP-was implemented, and from the DSRIP-period to after PRIME was implemented) in the average outcomes between demonstration HSAs and similar HSAs not influenced by DSRIP or PRIME.

We estimated an HSA-level regression model for an outcome (*y*) for HSA (*i*) at time (*t*):

$$
(2) \quad
\begin{aligned}
Y_{tij} = {} & \beta_0 + \beta_1(Post\_DSRIP_{ti}) + \beta_2(Demo_i) + \beta_3(Post\_DSRIP_{ti})*(Demo_i) + \\
& \beta_4(Post\_PRIME_{ti}) + \beta_5(Post\_PRIME_{ti})*(Demo_i) + \beta_6(W_{ti}) + \beta_7(Q_{ti}) + \varepsilon_{ti}
\end{aligned}
$$

This model includes three types of covariates:

- $Post\_DSRIP_{ti}$ is a post-demonstration period indicator, equal to 1 if the observation is in the post-DSRIP intervention period and equal to 0 if the observation is in the pre-intervention period.

- $Post\_PRIME_{ti}$ is a post-demonstration period indicator, equal to 1 if the observation is in the post-PRIME intervention period and equal to 0 if the observation is in the pre-intervention (DSRIP) period.

- $Demo_i$ is a demonstration indicator, equal to 1 if the HSA *i* is affected by the demonstration and equal to 0 if HSA *i* is in the comparison group.

- W_{ti} are HSA-level characteristics (such as percentage of beneficiaries that are children, percentage of beneficiaries with a chronic condition, whether the HSA is a primary care shortage area or a mental health shortage area).

- Q_{ti} are quarter indicators for three of four annual quarters to control for seasonal effects (for quarterly measures).

The model also includes a residual error term (ε_{ti}). The first coefficient of interest, β_3, represents the differential impact of DSRIP, relative to the comparison, on the outcome of interest. The second coefficient of interest, β_5, represents the differential impact of the PRIME relative to the DSRIP.

New York and Texas: simple interrupted time series

Unlike California, New York and Texas implemented DSRIP demonstrations statewide. As a result, these states did not have a large pool of in-state HSAs that we could use to identify a similar comparison group. Furthermore, potential out-of-state comparison groups with data available for our analysis differed from the intervention HSAs on key characteristics. Because we were unable to identify reasonable comparison groups in either New York or Texas, our analytic strategy relied on implementing a simple interrupted time series design in each state to estimate whether the level or trends in the outcomes of interest in the post-intervention period differed significantly from 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 in the absence of the intervention.

In New York, we estimated an HSA-level regression model for an outcome (y) for HSA (i) at time (t) using the following equation:

$$(3) \qquad Y_{ti} = \beta_0 + \beta_1(time_{ti}) + \beta_2(Post_i) + \beta_3(time_{ti}) * (Post_i) + \beta_4(W_{ti}) + \beta_5(Q_{ti}) + \varepsilon_{ti} ,$$

where:

- $Time_{ti}$ measures time since the start of the DSRIP demonstration (in years).

- $Post_i$ is a post-demonstration period indicator, equal to 1 if the observation is in the post-period and equal to 0 if the observation is in the pre-period.

- W_{ti} are characteristics of the HSA (such as the share of beneficiaries that are children, the share of residents with a behavioral health condition, the share of residents with one or more chronic conditions, the share of beneficiaries living in a primary care shortage area or a behavioral health shortage area, percentage of residents covered by Medicaid, hospital beds per capita), as well as quarter indicators to control for seasonal effects.

- Q_{ti} are quarter indicators for three of four annual quarters to control for seasonal effects (for quarterly measures).

The model also includes a residual error term (ε_{ti}).

In Texas, we estimated individual-level regression models for an outcome (y) for a beneficiary (i) at time (t) using the following equation[22]:

$$(4) \quad Y_{ti} = \beta_0 + \beta_1(time_{ti}) + \beta_2(Post_i) + \beta_3(time_{ti}) * (Post_i) + \beta_4(W_i) + \beta_5(Q_{ti}) + \varepsilon_{ti} ,$$

---

[22] Texas Medicaid data had significant data quality concerns relating to zip codes, which precluded analysis at the HSA level.

where:

- *Time*$_{ti}$ measures time since the start of the DSRIP demonstration (in years).

- *Post*$_i$ is a post-demonstration period indicator, equal to 1 if the observation is in the post-period and equal to 0 if the observation is in the pre-period.

- *W*$_{ti}$ are characteristics of the individual (such as age, sex, presence of a behavioral health condition, and presence of one or more chronic conditions).

- *Q*$_{ti}$ are quarter indicators for three of four annual quarters to control for seasonal effects (for quarterly measures).

The model also includes a residual error term (ε_{ti}).

The coefficients of interest in both models, β_2 and β_3, represent the change on the *level* and the (linear) *trend* of the outcome variable between the pre- and post-intervention periods. In particular, a statistically significant and substantively large β_2 would indicate that an immediate change in outcome occurred at the time of the DSRIP implementation (when time is centered at the point of DSRIP implementation), whereas a statistically significant and substantively large β_3 would indicate that the annual or quarterly rate of change in the outcome variable differs in the pre- and post-demonstration periods.

B. Outcomes derived from Healthcare Cost and Utilization Project data

New Jersey and Texas: Simple interrupted time series

Because we only had access to Healthcare Cost and Utilization Project data in New Jersey and Texas, our analytic strategy relied on implementing a simple interrupted time series design in each state to estimate whether the trends in hospital discharges for ACSCs measured in the post-intervention period differed significantly from what would be expected in the absence of the intervention.

We estimated an HSA-level regression model for the outcome (*y),* the number of individuals in the HSA with discharges for selected chronic conditions per 100,000 Medicaid beneficiaries and uninsured individuals in an HSA (*i*), at time (*t),* using the following equation:

(5) $Y_{ti} = \beta_0 + \beta_1(time_{ti}) + \beta_2(Post_i) + \beta_3(time_{ti}) * (Post_i) + \beta_4(Q_{ti}) + \alpha_i + \varepsilon_{ti}$,

where:

- *Time*$_{ti}$ measures time since the start of the DSRIP demonstration (in quarters).

- *Post*$_i$ is a post-demonstration period indicator, equal to 1 if the observation is in the post-period and equal to 0 if the observation is in the pre-period.

- Q_{ti} are quarter indicators for three of four annual quarters to control for seasonal effects.

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

The coefficients of interest, β_2 and β_3, represent the change in the *level* and the (linear) *trend* of the outcome variable between the pre- to post-intervention periods. A statistically significant and substantively large $\widehat{\beta}_2$ would indicate that an immediate change in the measure occurred at the time of the DSRIP implementation (when time is centered at the point of DSRIP implementation), whereas a statistically significant and substantively large $\widehat{\beta}_3$ would indicate that the annual rate of change in the outcome variable differed in the pre- and post-demonstration periods.

We estimated these models overall, and then we estimated separate models for asthma-related conditions, diabetes-related conditions, and cardiovascular-related conditions.

Appendix B:

Unadjusted Trends

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

California

Figure B.1. ED visits per 1,000 beneficiaries in California: demonstration and comparison groups by quarter, January 2009 through December 2017
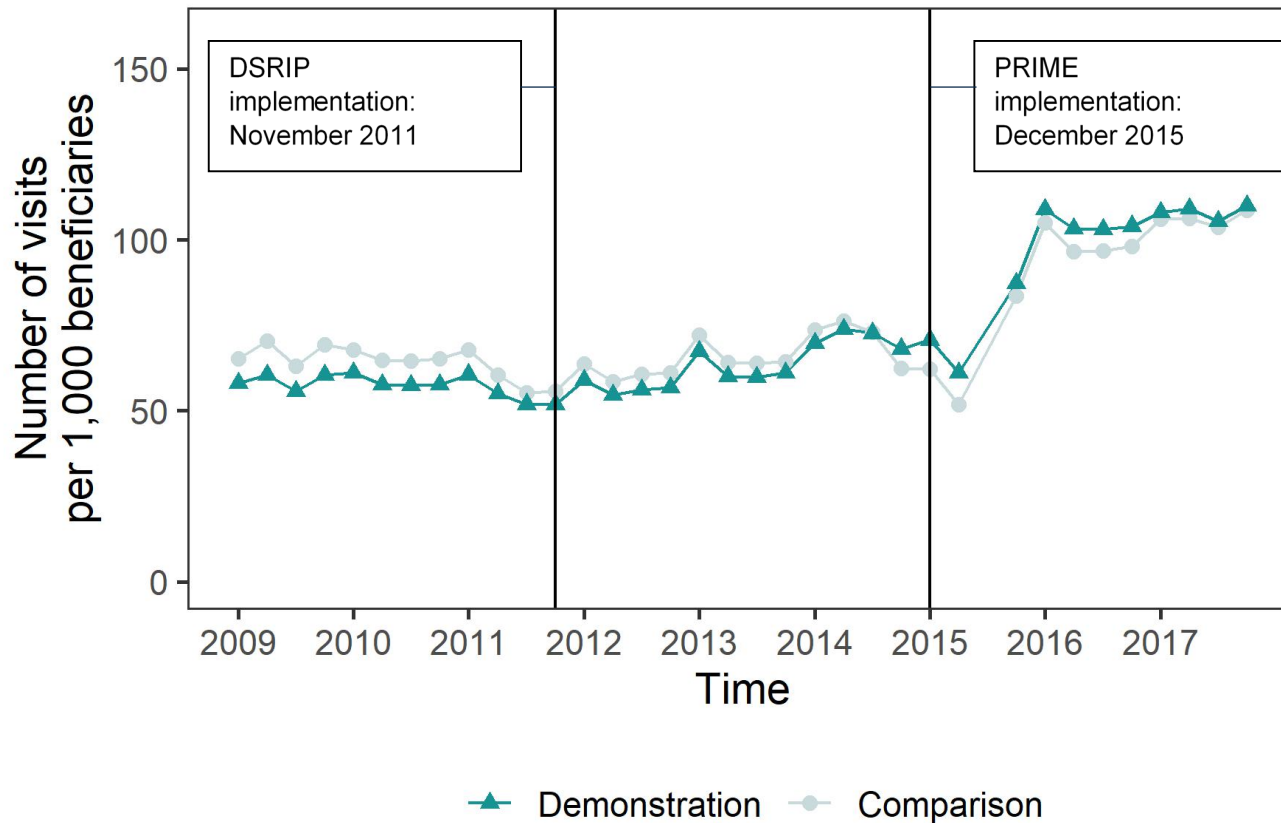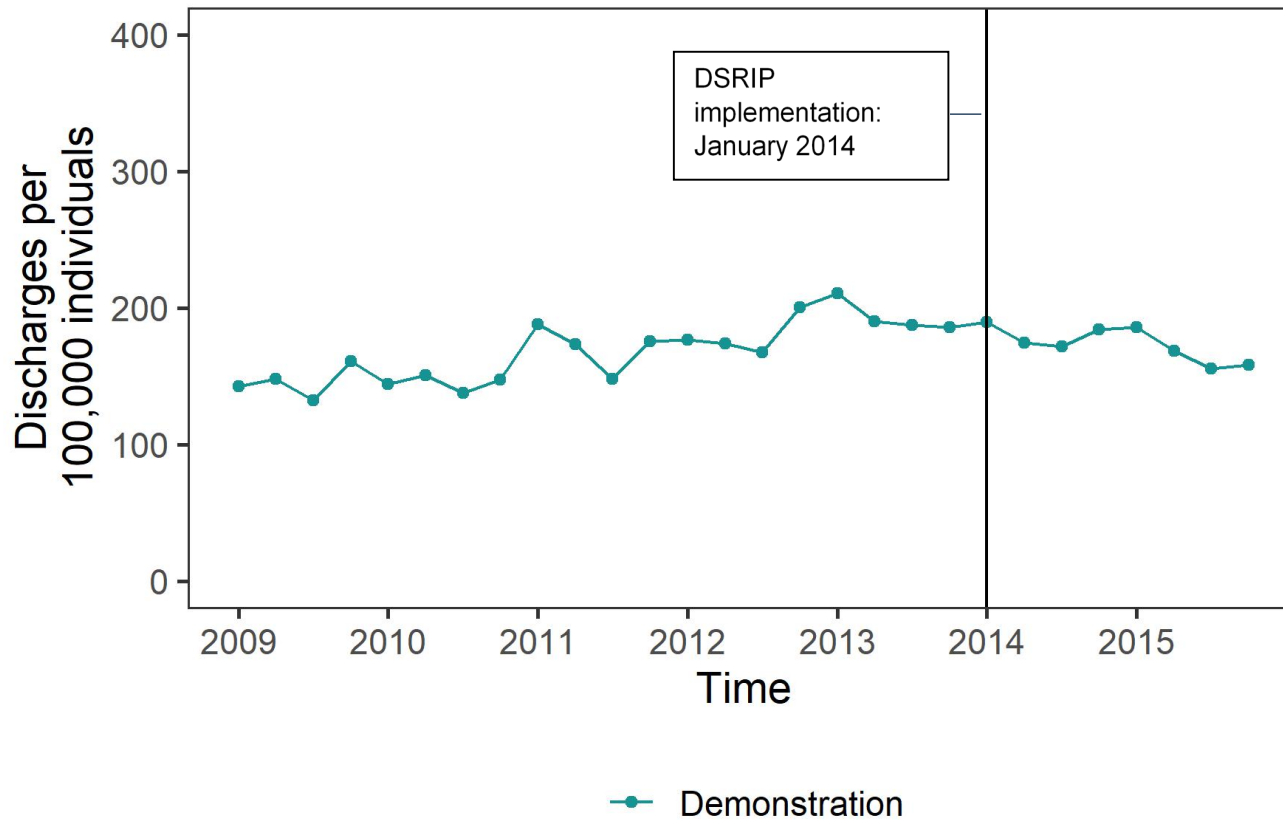

Source: Mathematica's analysis of TAF data from California, Virginia, and Washington from 2009 through 2017.

Note: We dropped data from 2015 Q3, the quarter before the state transitioned to reporting Medicaid data through T-MSIS.

DSRIP = delivery system reform incentive payment; PRIME = Public Hospital Redesign in Medi-Cal.

Figure B.2. Avoidable ED visits per 1,000 beneficiaries in California: demonstration and comparison groups by quarter, January 2009 through December 2017
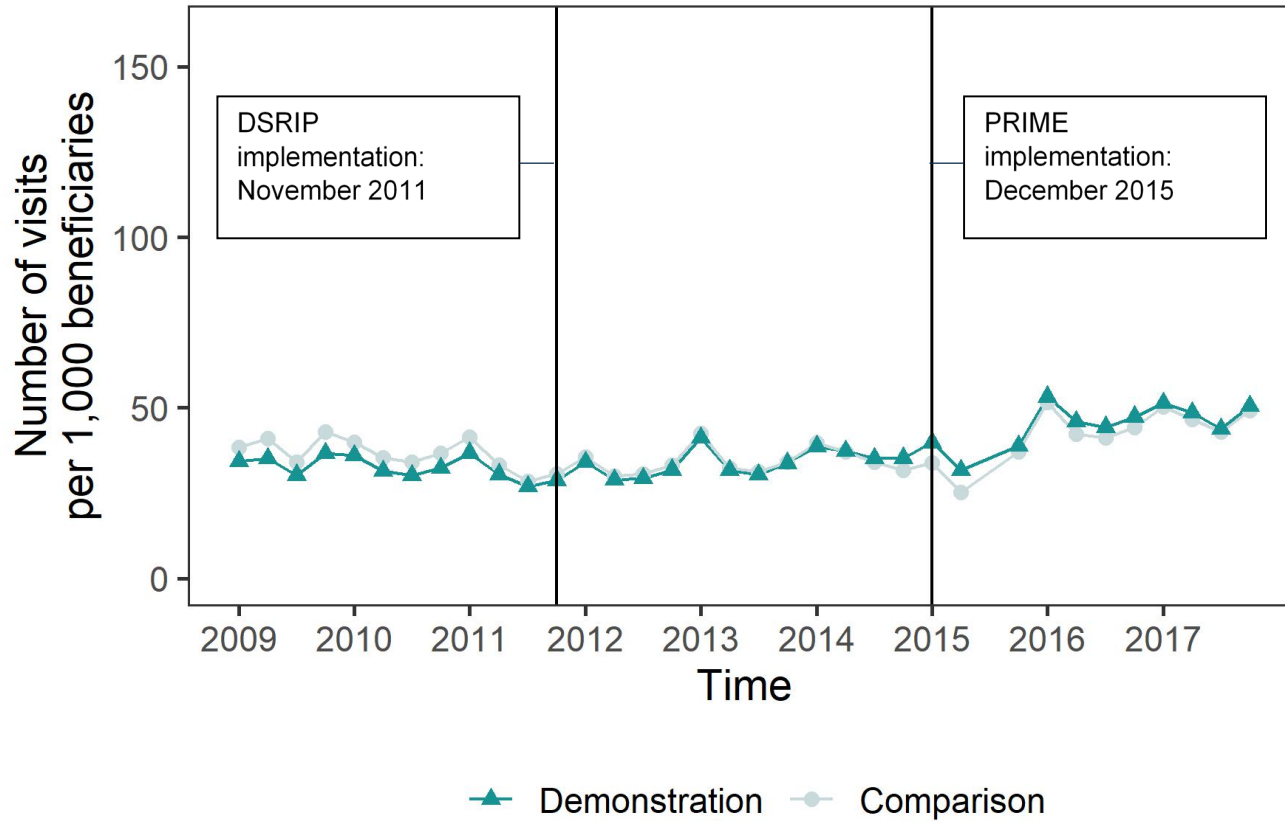
Source: Mathematica's analysis of TAF data from California, Virginia, and Washington from 2009 through 2017.

Note: We dropped data from 2015 Q3, the quarter before the state transitioned to reporting Medicaid data through T-MSIS.

DSRIP = delivery system reform incentive payment; PRIME = Public Hospital Redesign in Medi-Cal.

Figure B.3. Percentage of adult beneficiaries with an ambulatory care visit in California: demonstration and comparison groups by year, 2009 through 2017
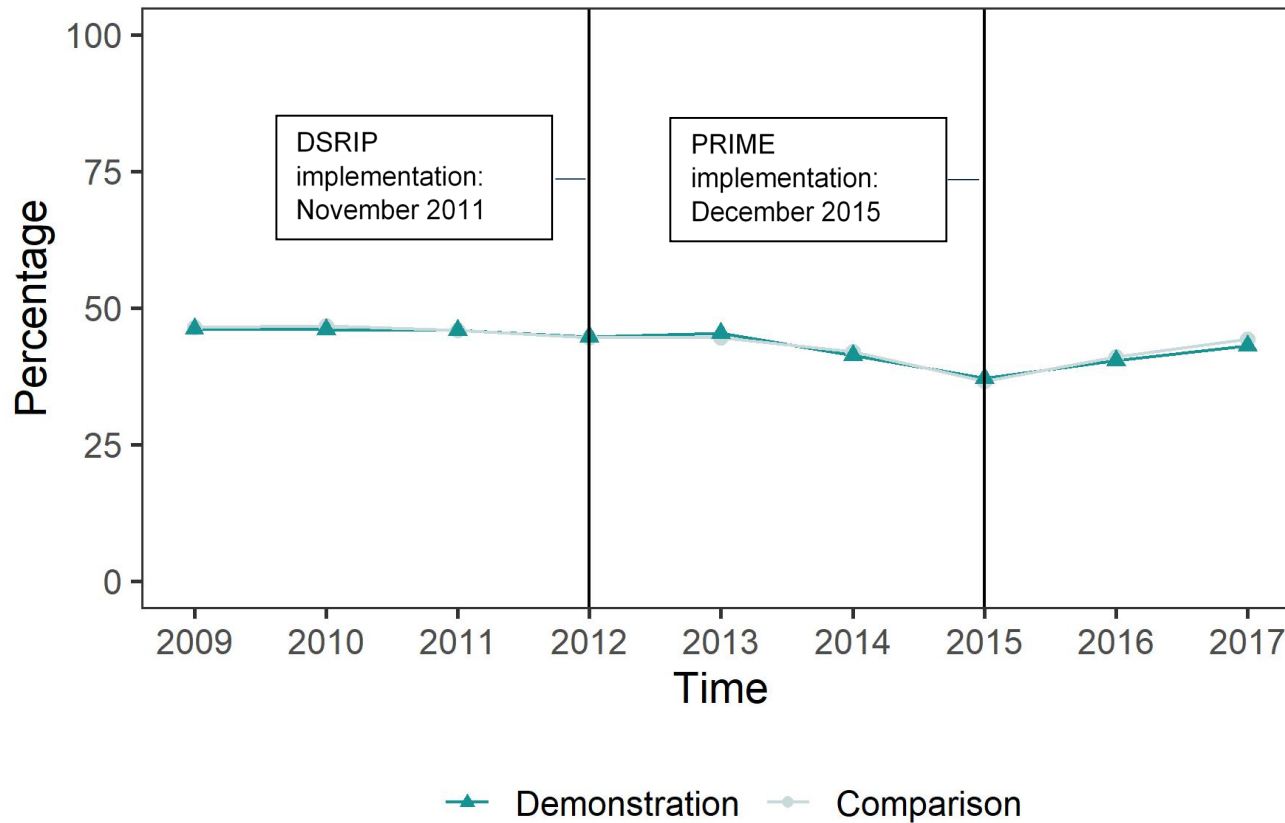
DSRIP = delivery system reform incentive payment; PRIME = Public Hospital Redesign in Medi-Cal.

Figure B.4. Percentage of child and adolescent beneficiaries with a primary care visit in California: demonstration and comparison groups by year, 2009 through 2017
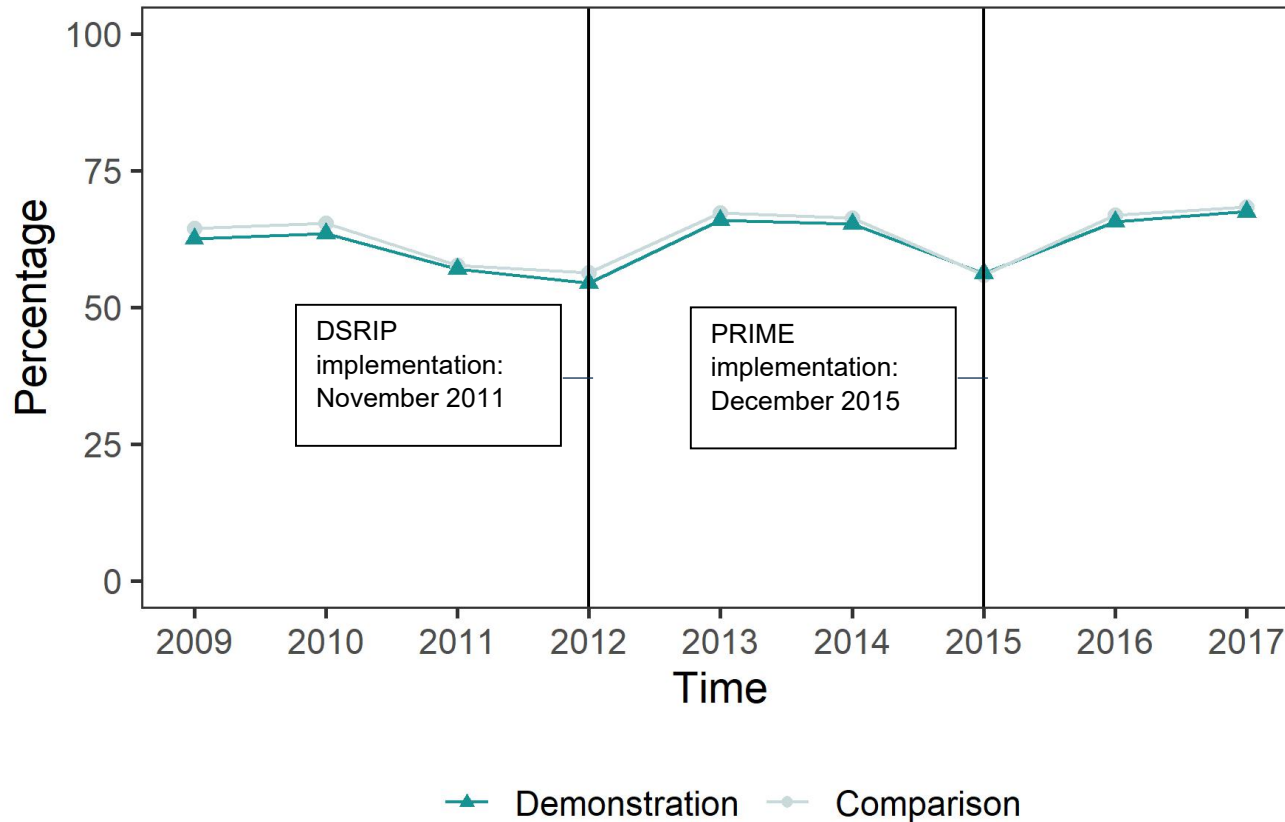
DSRIP = delivery system reform incentive payment; PRIME = Public Hospital Redesign in Medi-Cal.

Figure B.5. Percentage of beneficiaries with a behavioral health visit in California: demonstration and comparison groups by year, 2009 through 2017
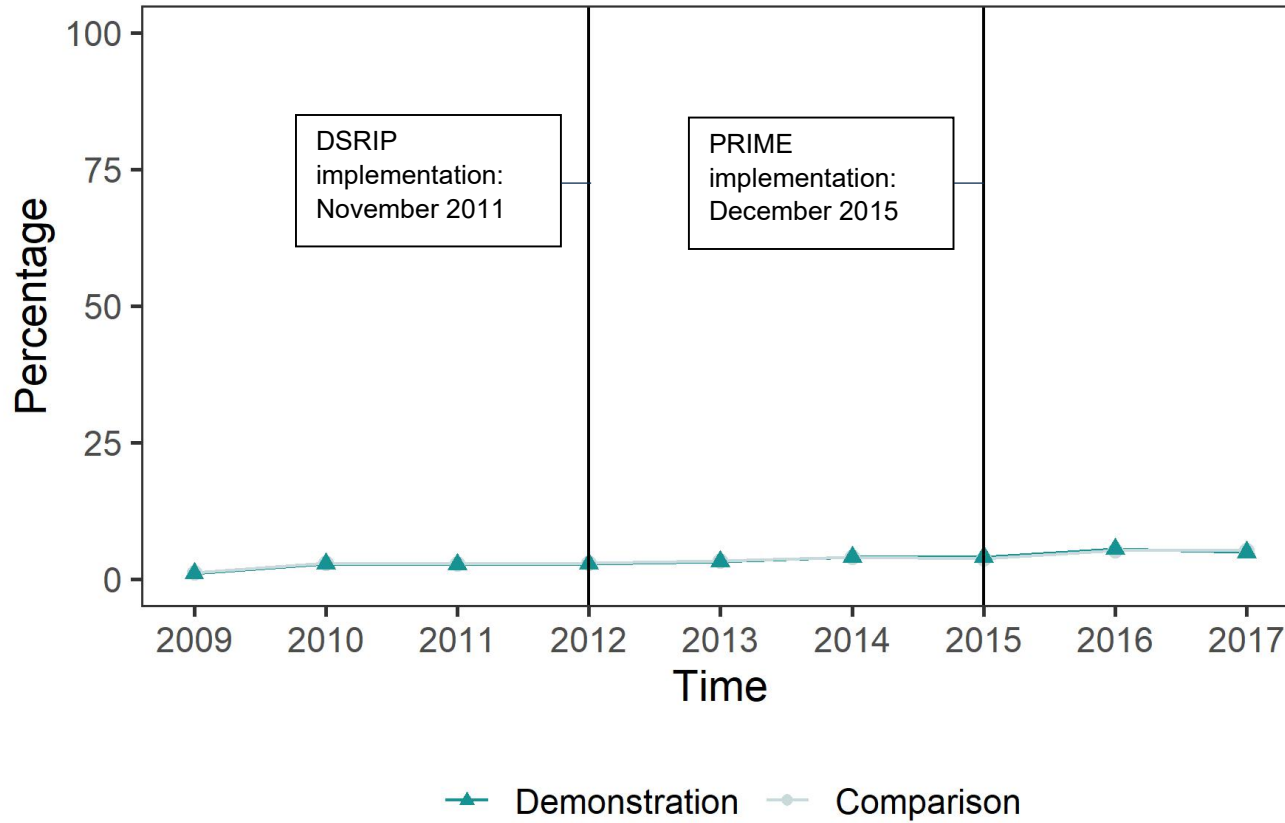
Figure B.6. Percentage of beneficiaries with diabetes who had an HbA1c test in California: demonstration and comparison groups by year, 2009 through 2017
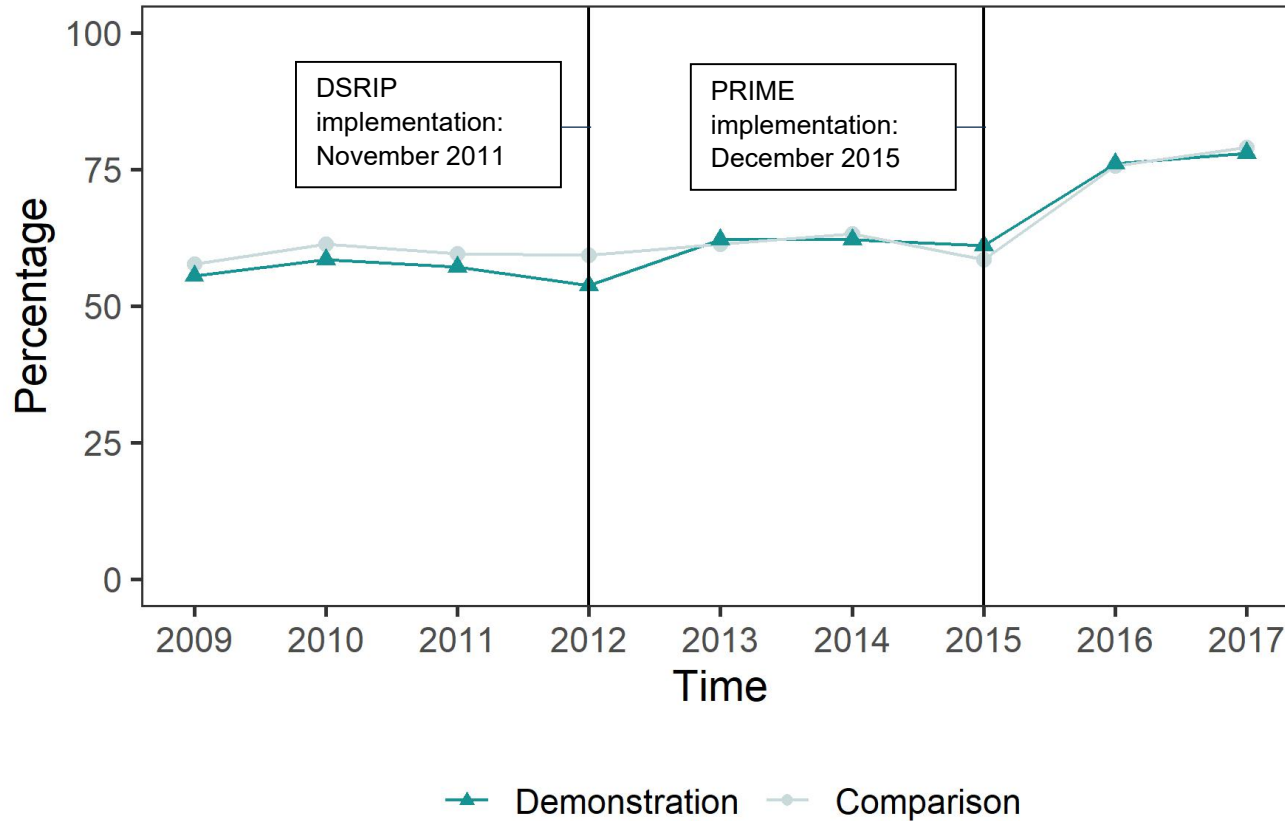
Source: Mathematica's analysis of TAF data from California, Virginia, and Washington from 2009 through 2017.

DSRIP = delivery system reform incentive payment; PRIME = Public Hospital Redesign in Medi-Cal.

Figure B.7. Percentage of ED visits for ACSCs with a follow-up within seven days in California: demonstration and comparison groups by quarter, January 2009 through December 2017
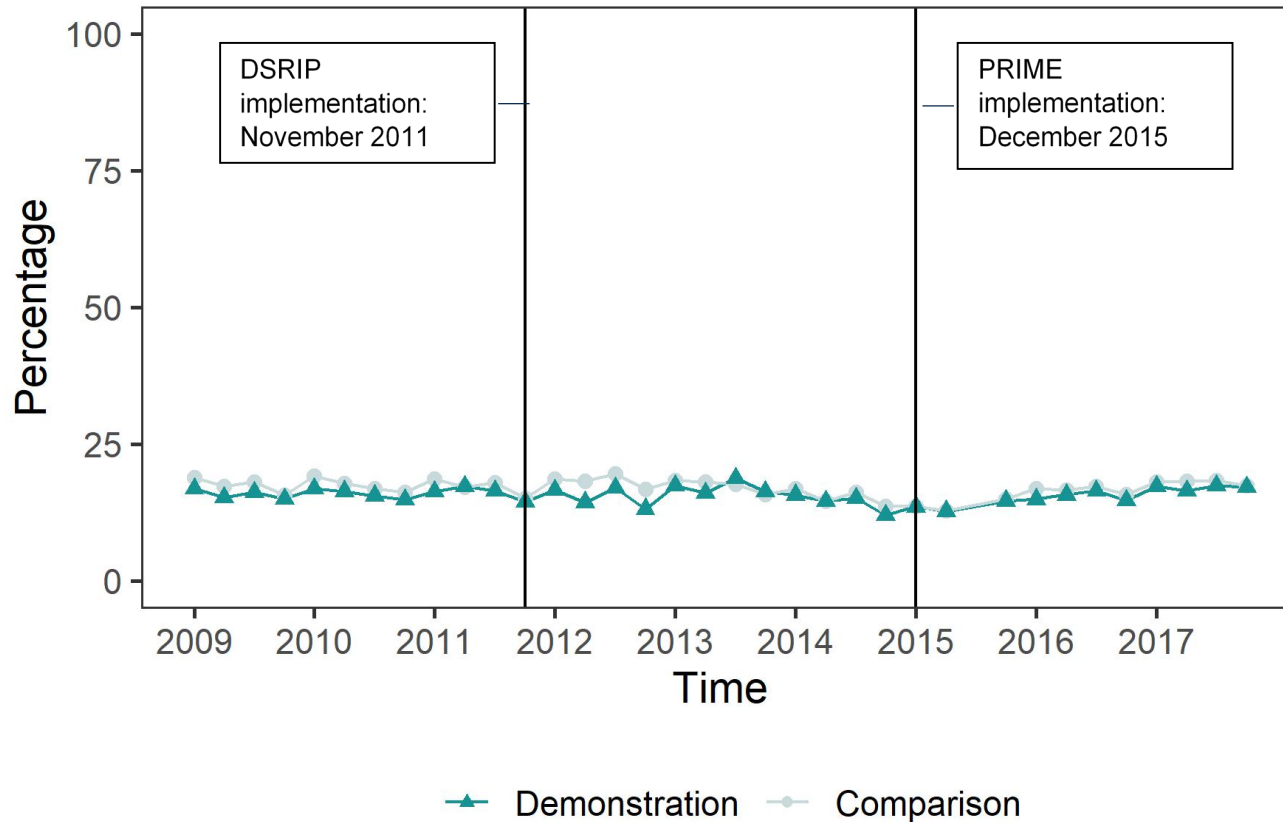
New Jersey

Figure B.8. Hospital discharges for ACSCs per 100,000 Medicaid beneficiaries and uninsured individuals by quarter in New Jersey, January 2009 through December 2015
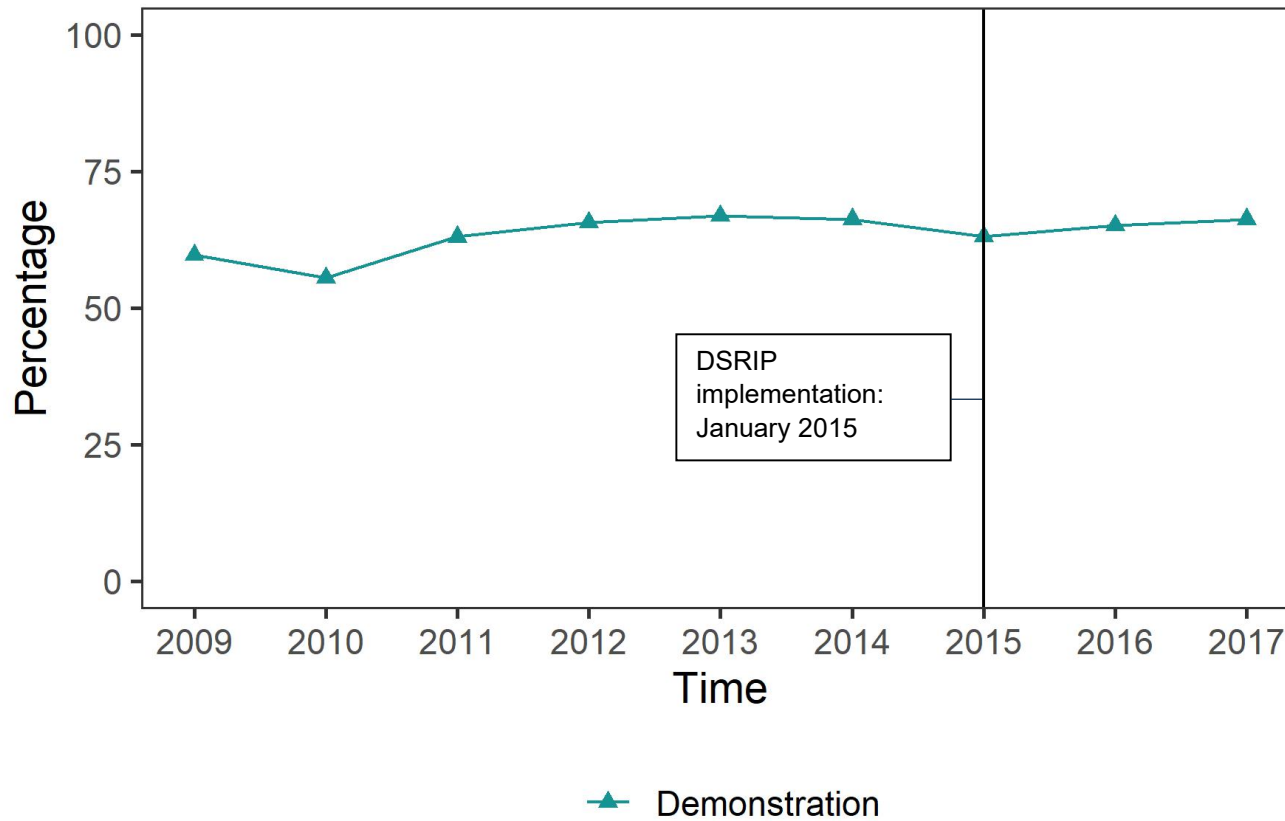

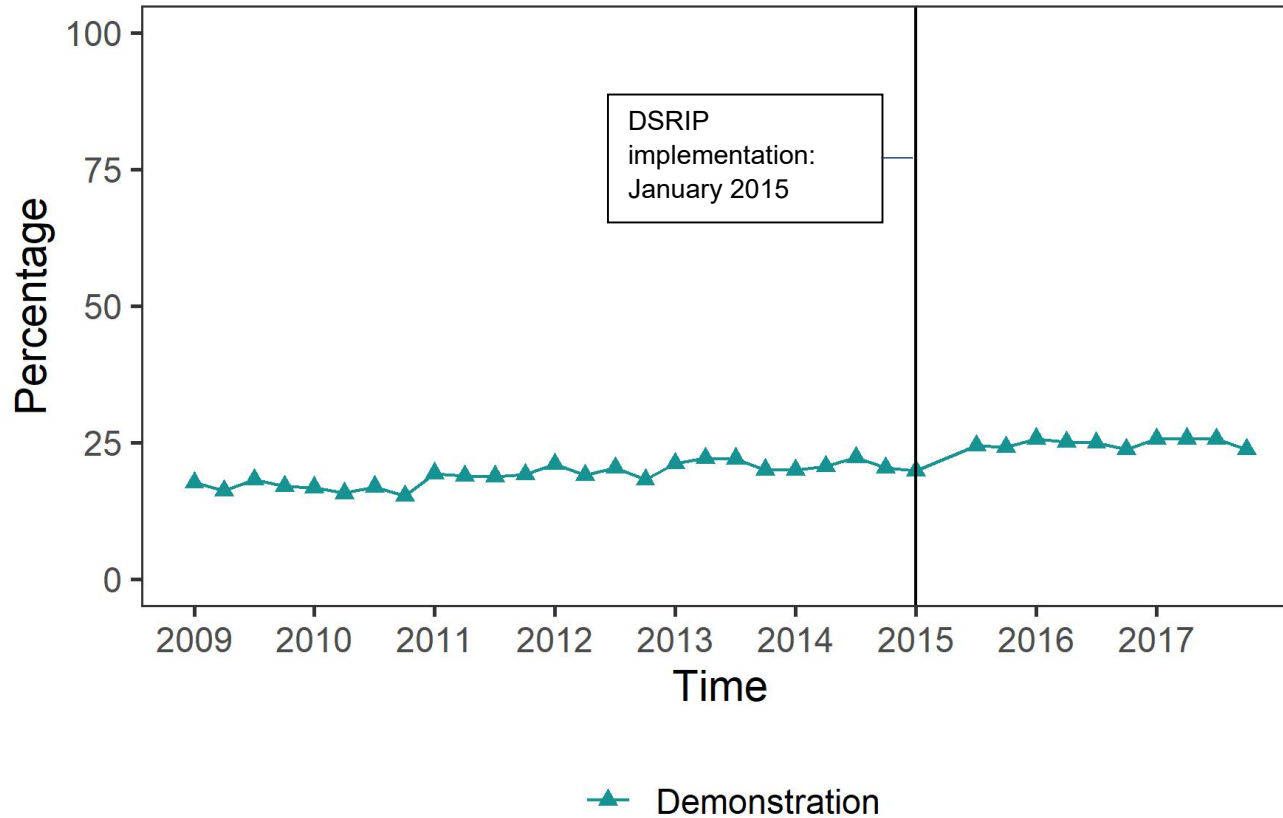Source: Mathematica's analysis of HCUP data from New Jersey, 2009 to 2015.

DSRIP = delivery system reform incentive payment.

Figure B.9. Hospital discharges for ACSCs per 100,000 individuals by condition category, by quarter in New Jersey, January 2009 through December 2015

New York

Figure B.10. ED visits per 1,000 beneficiaries by quarter in New York, January 2009 through December 2017
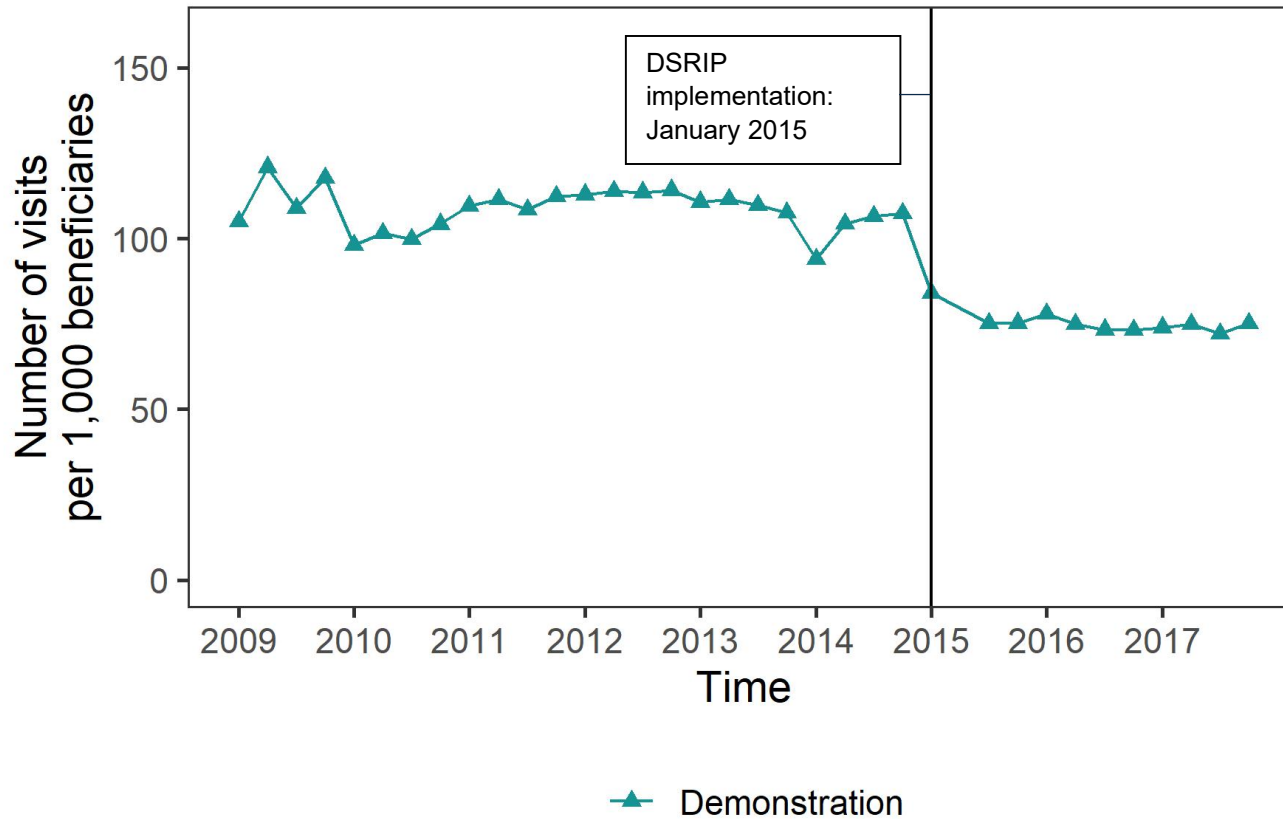
Source: Mathematica's analysis of TAF data from New York from 2009 through 2017.

Note: We dropped data from 2015 Q2, the quarter before the state switched to reporting Medicaid data through the T-MSIS. DSRIP = delivery system reform incentive payment.

Figure B.11. Avoidable ED visits per 1,000 beneficiaries by quarter in New York, January 2009 through December 2017
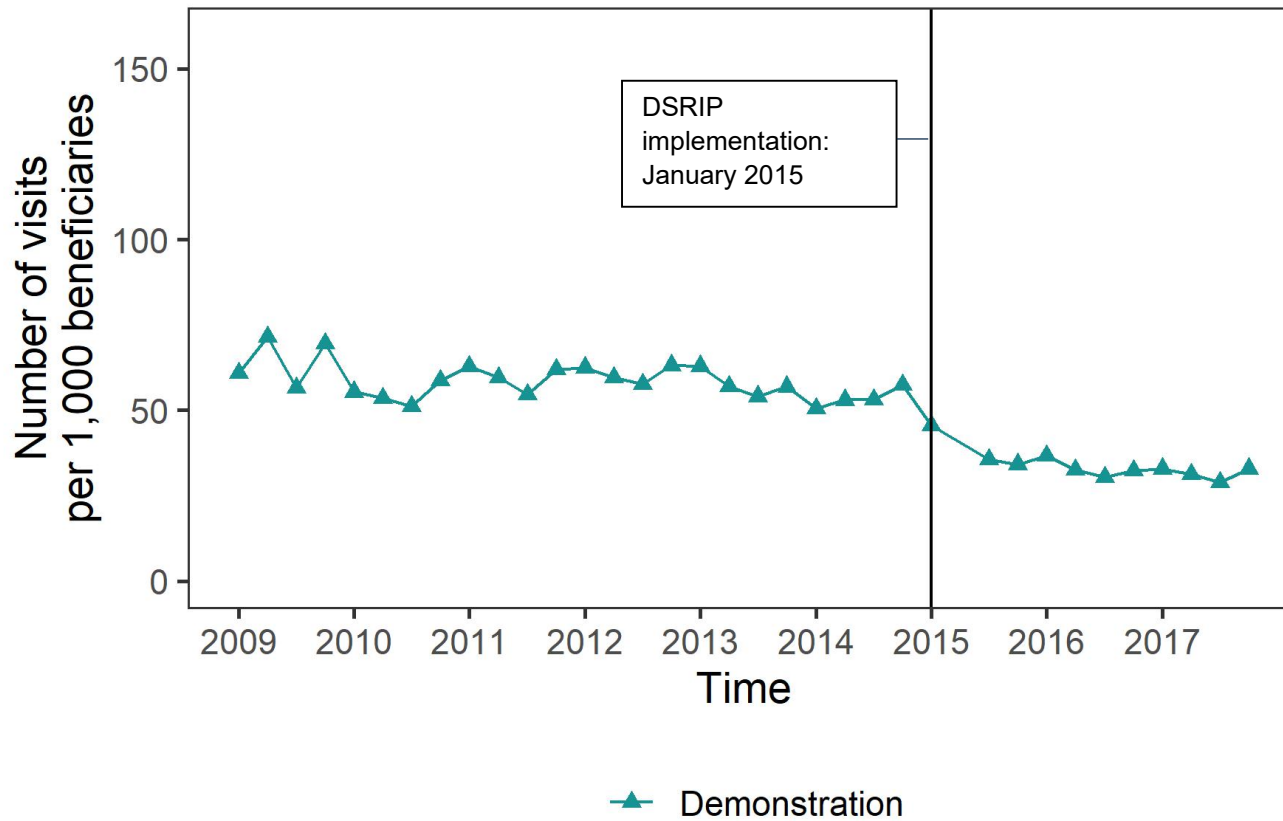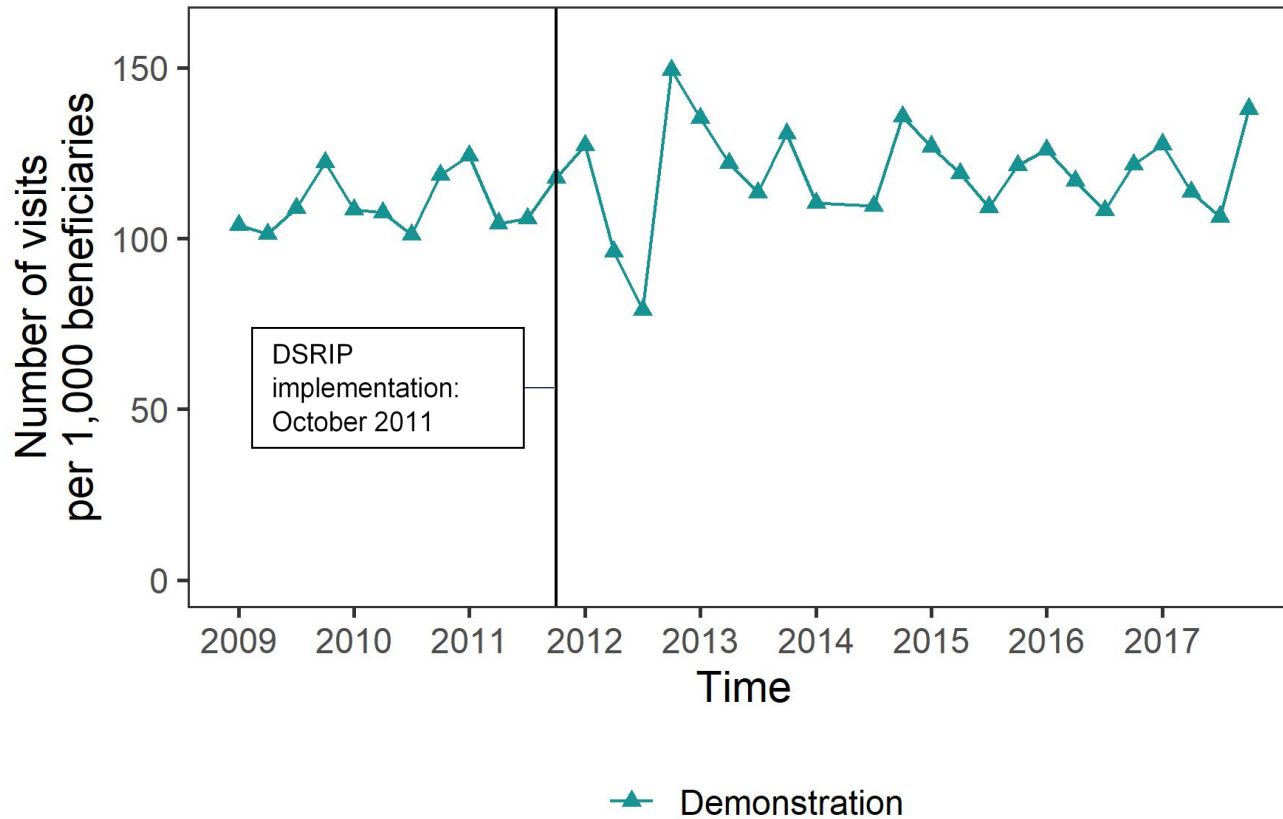

Source: Mathematica's analysis of TAF data from New York from 2009 through 2017.

Note: We dropped data from 2015 Q2, the quarter before the state switched to reporting Medicaid data through the T-MSIS.

DSRIP = delivery system reform incentive payment.

Figure B.12. Percentage of adult beneficiaries with an ambulatory care visit by year in New York, 2009 through 2017
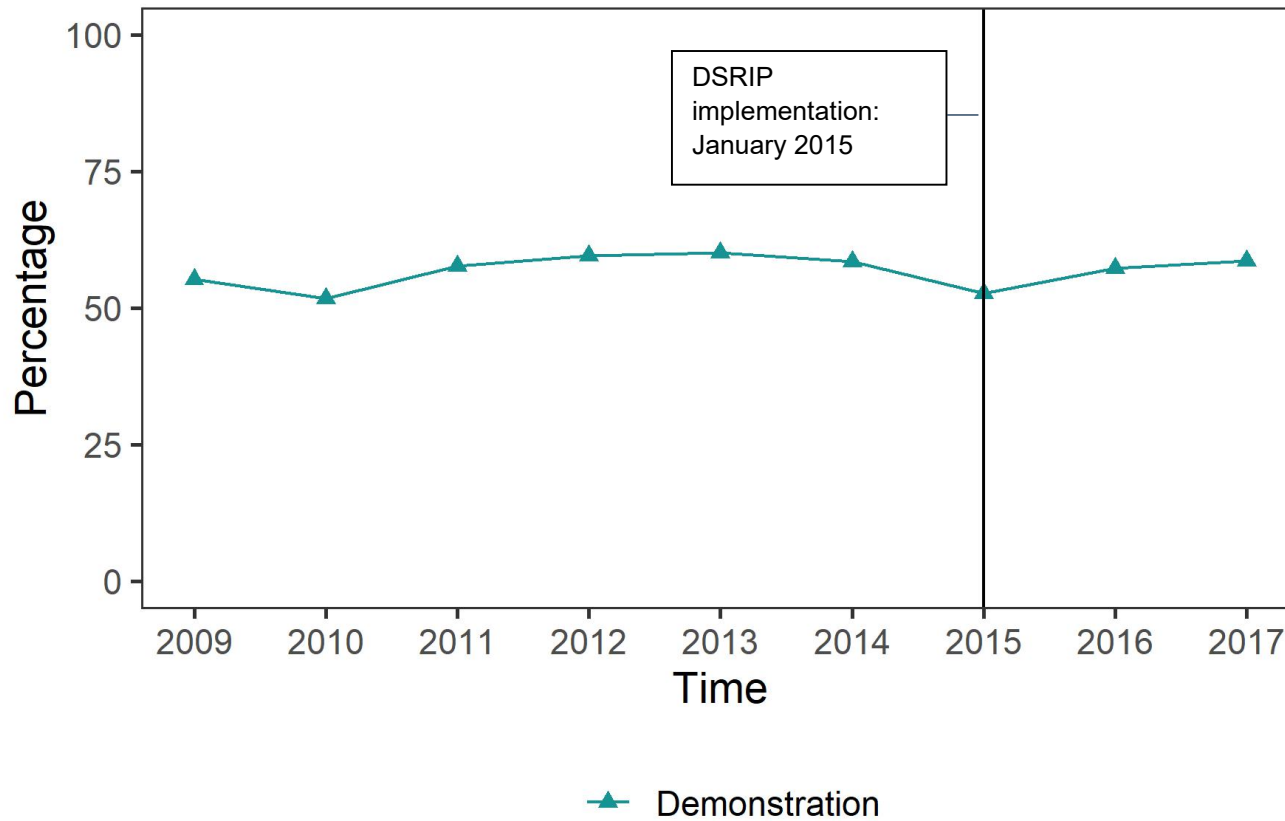
Source: Mathematica's analysis of TAF data from New York from 2009 through 2017.
DSRIP = delivery system reform incentive payment.

Figure B.13. Percentage of child and adolescent beneficiaries with a primary care visit by year in New York, 2009 through 2017
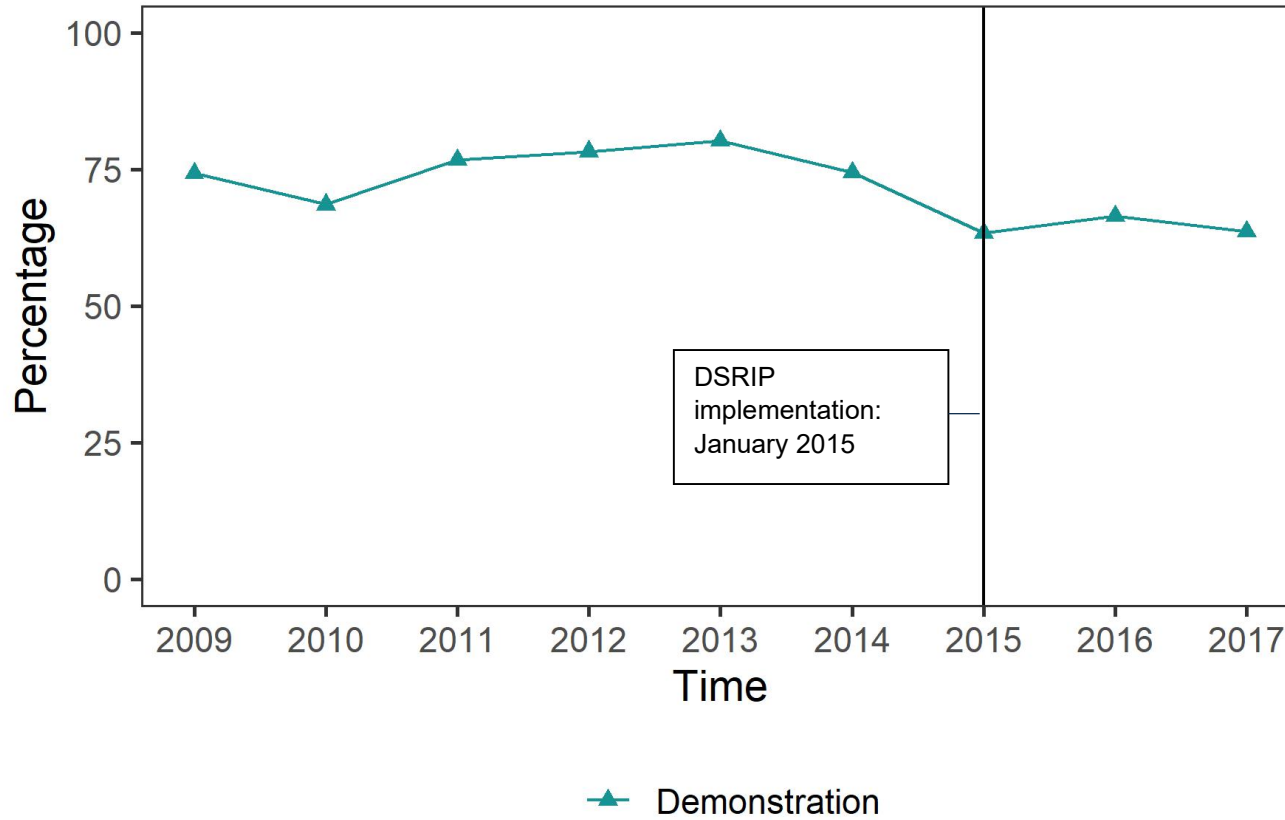
Source: Mathematica's analysis of TAF data from New York from 2009 through 2017.

DSRIP = delivery system reform incentive payment.

Figure B.14. Percentage of beneficiaries with a behavioral health visit by year in New York, 2009 through 2017
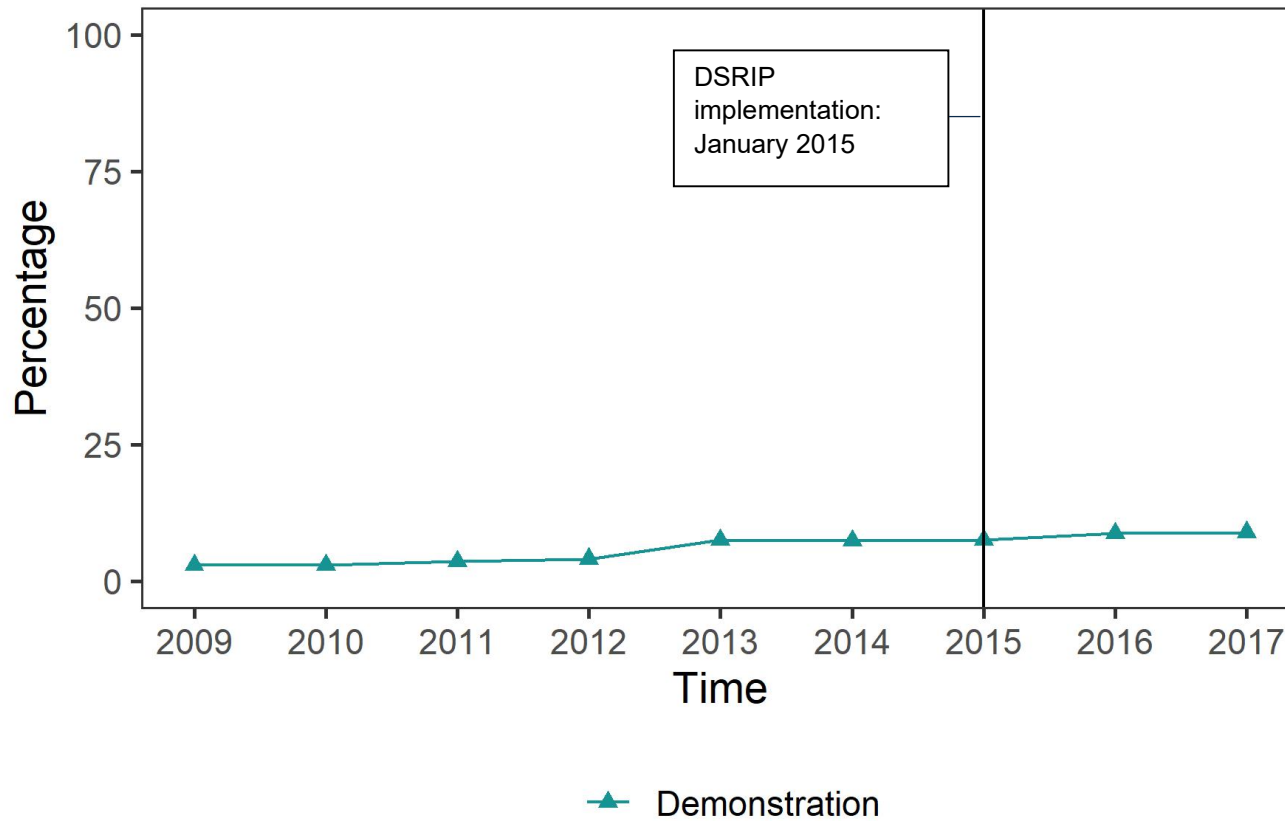
Source: Mathematica's analysis of TAF data from New York from 2009 through 2017.

DSRIP = delivery system reform incentive payment.

Figure B.15. Percentage of beneficiaries with diabetes who had an HbA1c test by year in New York, 2009 through 2017

Figure B.16. Percentage of ED visits for ACSCs with a follow-up visit within seven days by quarter in New York, January 2009 through December 2017
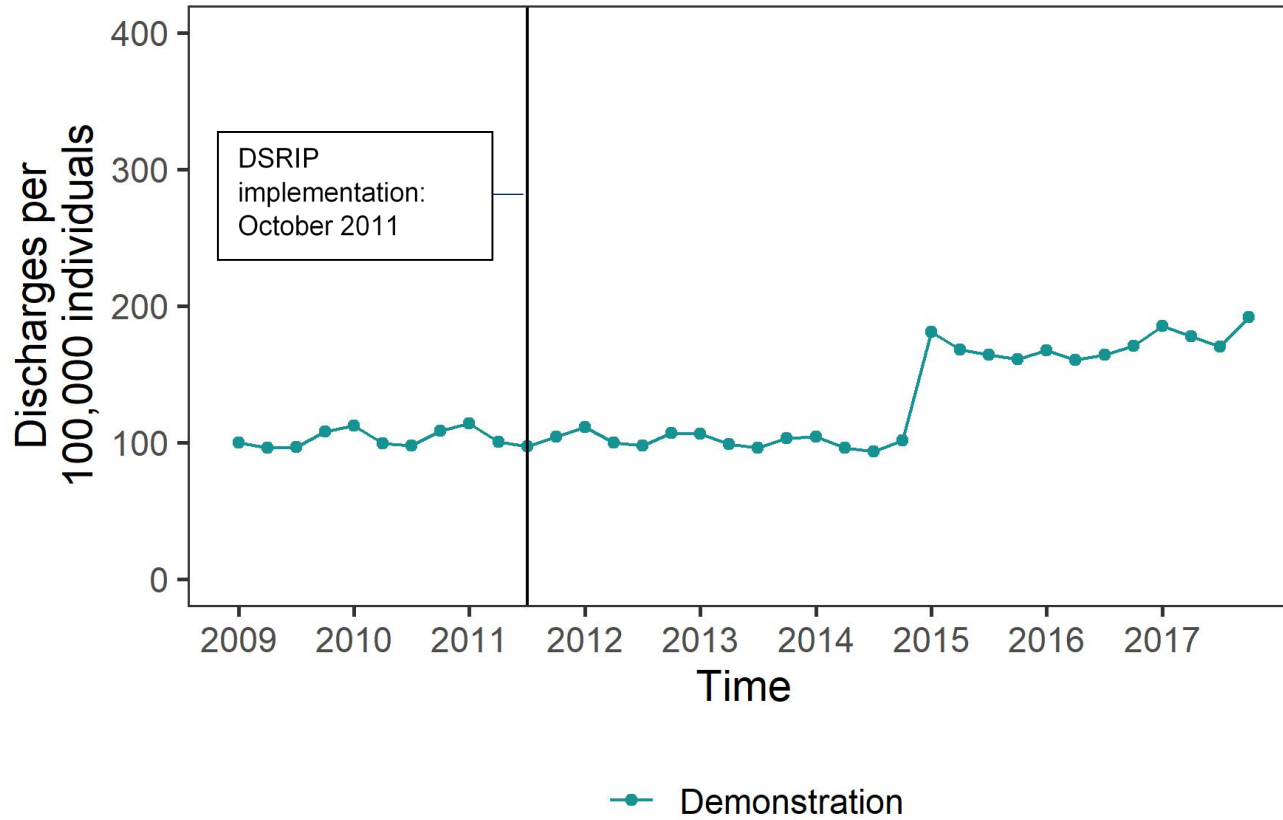

Source: Mathematica's analysis of TAF data from New York from 2009 through 2017.

Note: We dropped data from 2015 Q2, the quarter before the state transitioned to reporting Medicaid data through T-MSIS.

DSRIP = delivery system reform incentive payment.

Texas

Figure B.17. ED visits per 1,000 beneficiaries by quarter in Texas, January 2009 through December 2017

Source: Mathematica's analysis of TAF data from Texas from 2009 through 2017.

Note: We dropped data from 2013 Q4, the quarter before the state transitioned to reporting Medicaid data through T-MSIS.
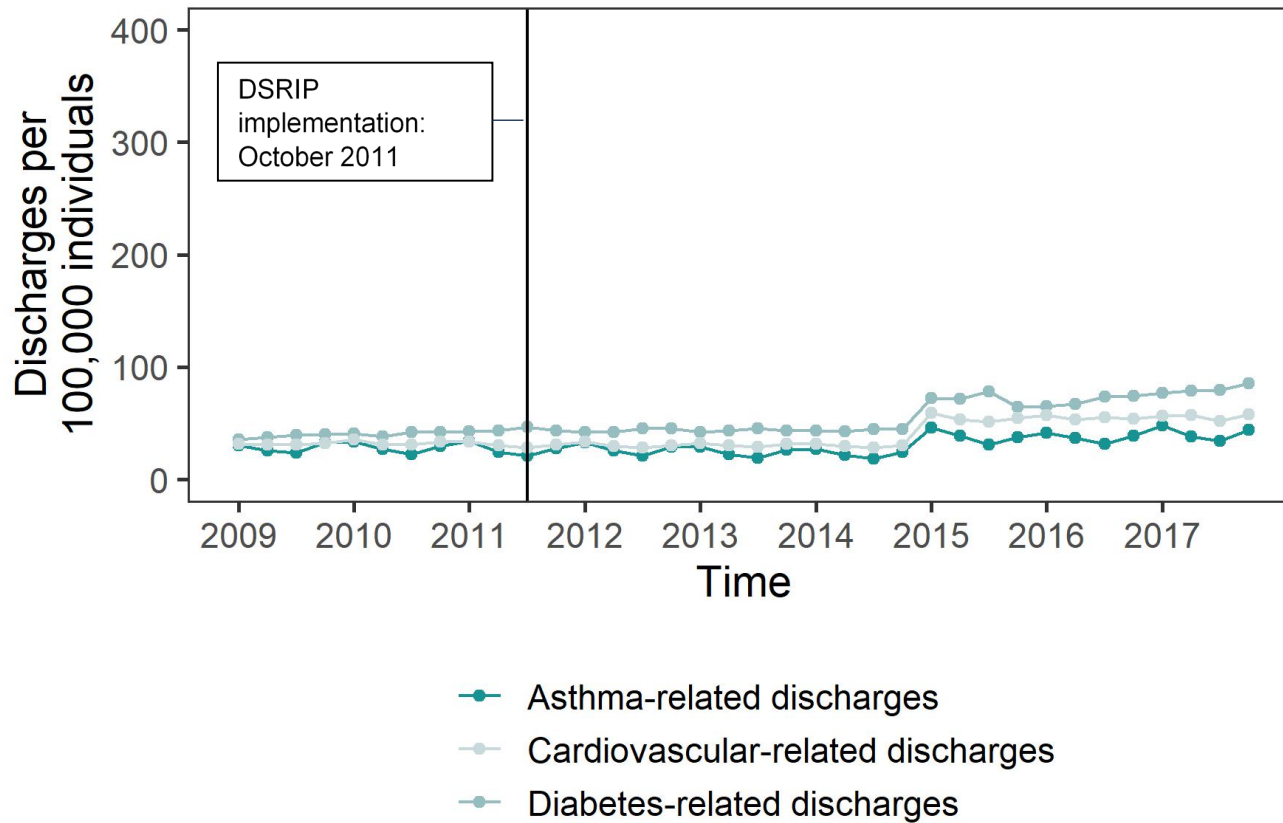
DSRIP = delivery system reform incentive payment.

Figure B.18. Avoidable ED visits per 1,000 beneficiaries by quarter in Texas, January 2009 through December 2017
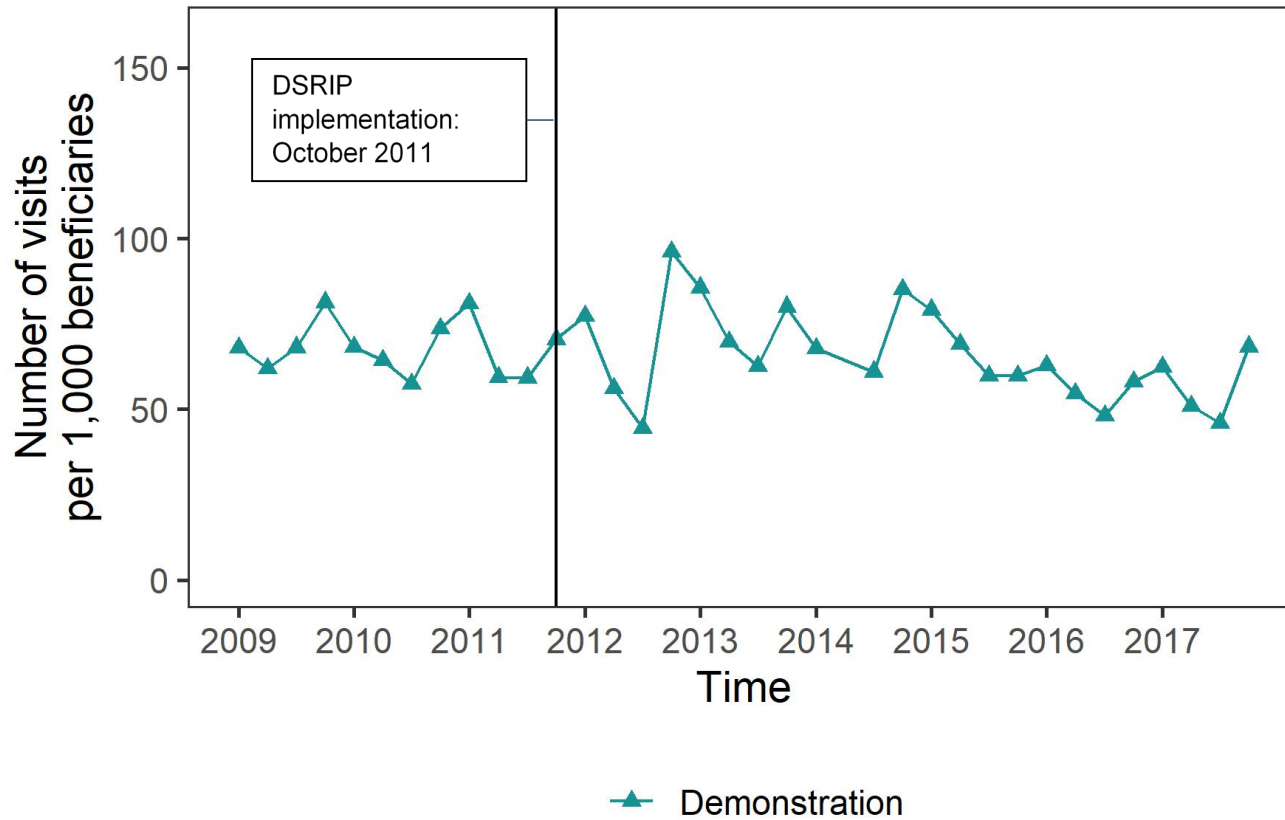
Source: Mathematica's analysis of TAF data from Texas from 2009 through 2017.

Note: We dropped data from 2013 Q4, the quarter before the state switched to reporting Medicaid data through the T-MSIS.

DSRIP = delivery system reform incentive payment.

Figure B.19. Hospital discharges for ACSCs per 100,000 Medicaid beneficiaries and uninsured individuals by quarter in Texas, January 2009 through December 2015
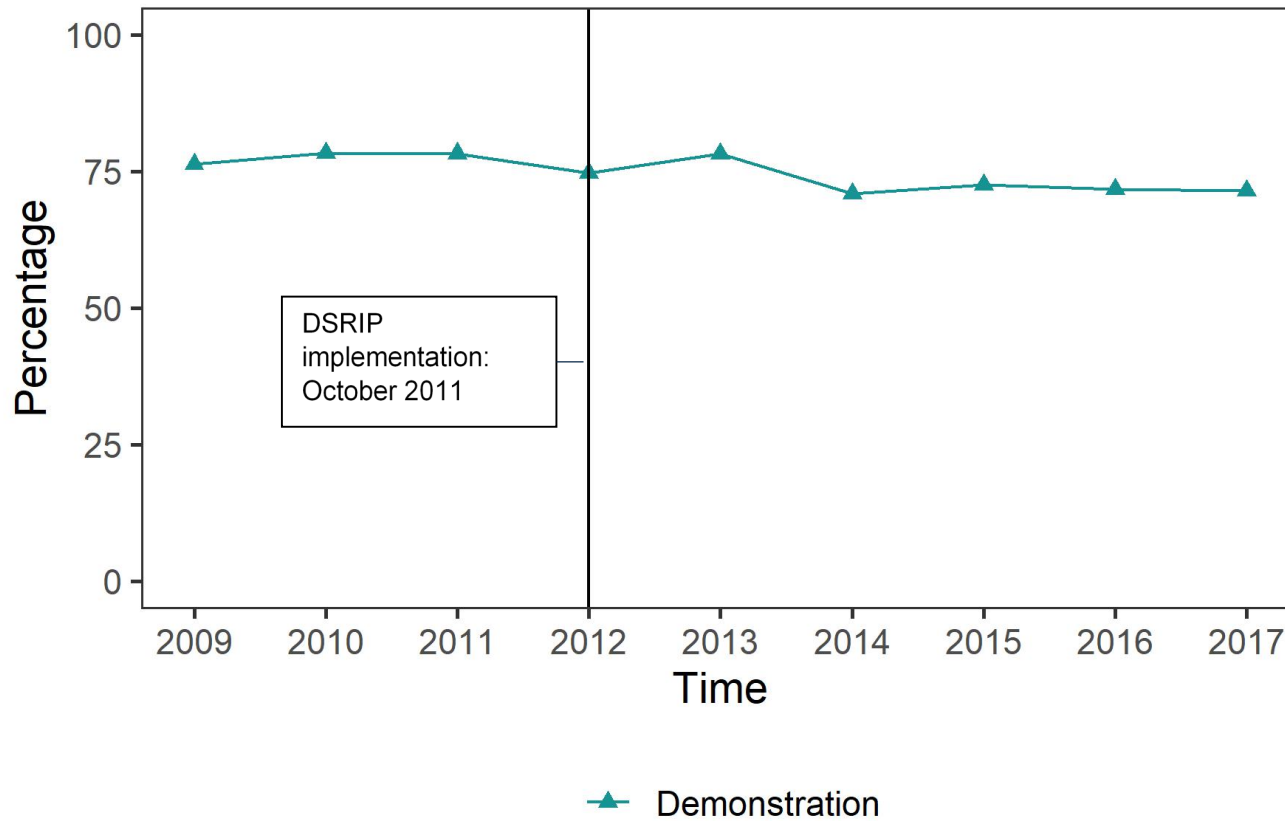

Source: Mathematica's analysis of HCUP data from Texas, 2009 to 2017.

DSRIP = delivery system reform incentive payment.

Figure B.20. Hospital discharges for ACSCs per 100,000 Medicaid beneficiaries and uninsured individuals by condition category, by quarter in Texas, January 2009 through December 2017

Figure B.21. Percentage of adult beneficiaries with an ambulatory care visit by year in Texas, 2009 through 2017
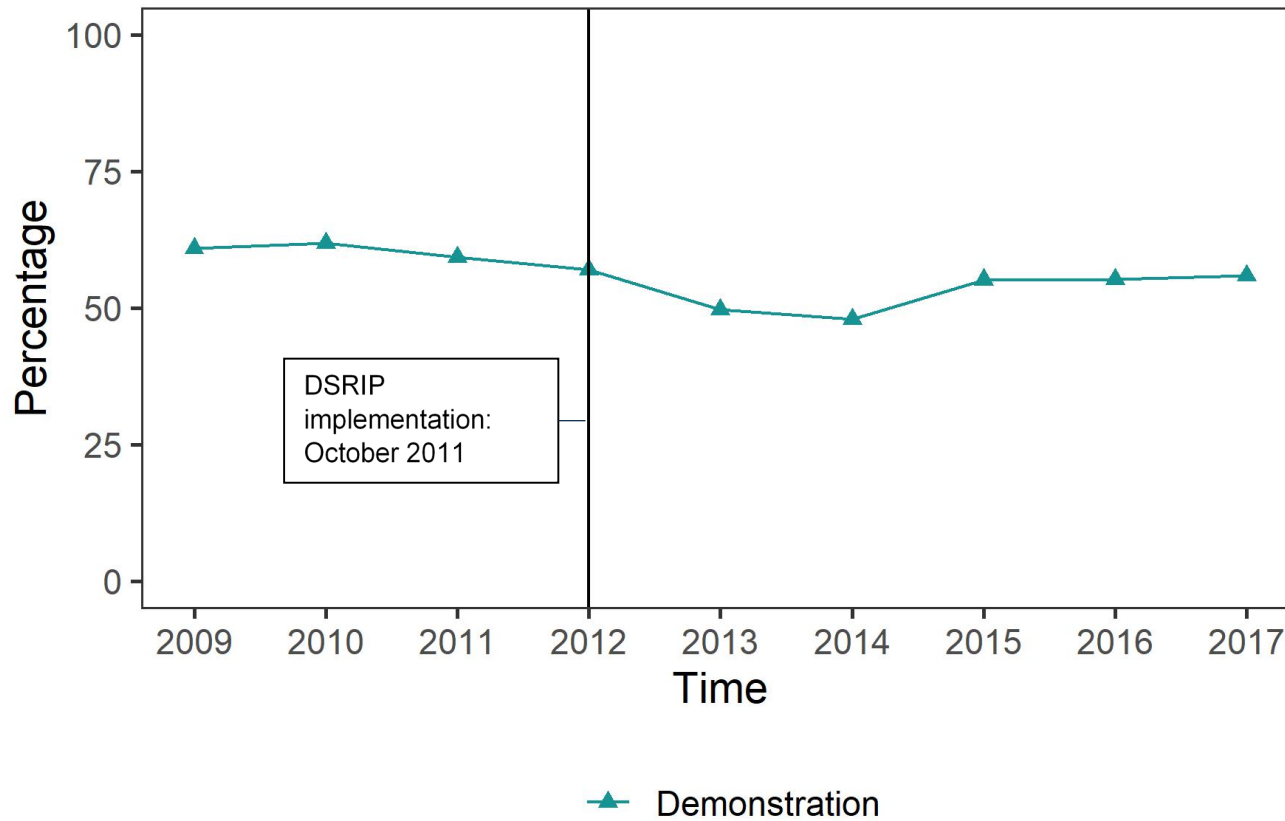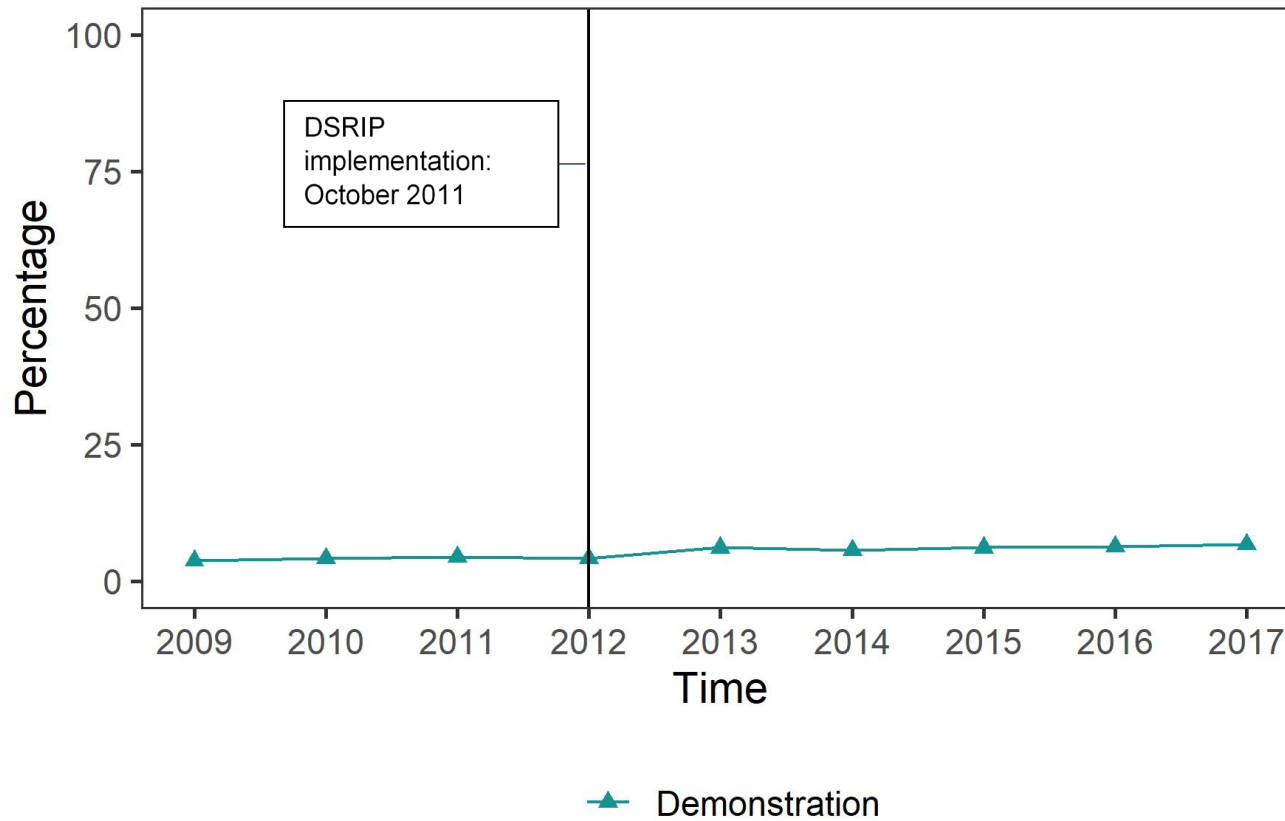

Source: Mathematica's analysis of TAF data from Texas from 2009 through 2017.

DSRIP = delivery system reform incentive payment.

Figure B.22. Percentage of child and adolescent beneficiaries with a primary care visit by year in Texas, 2009 through 2017
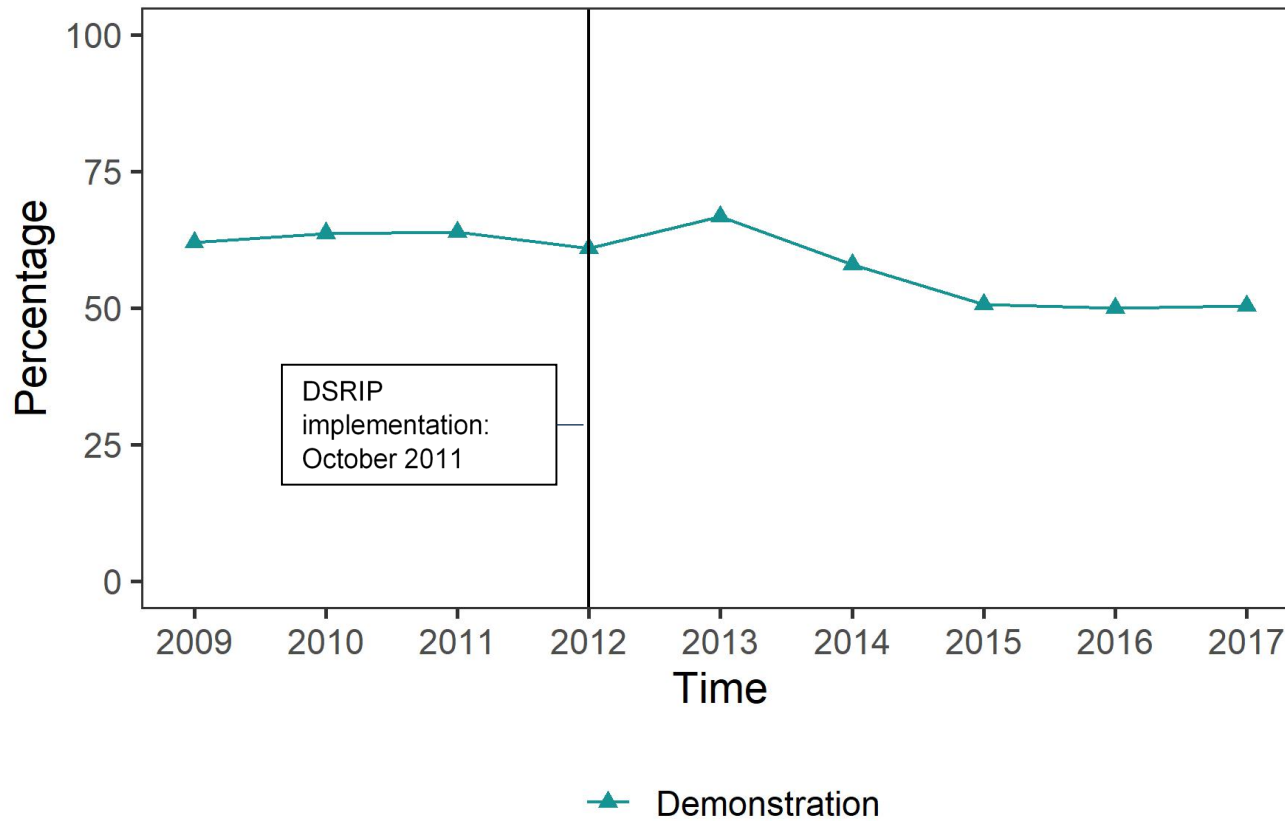

Source: Mathematica's analysis of TAF data from Texas from 2009 through 2017.
DSRIP = delivery system reform incentive payment.

**Figure B.23. Percentage of beneficiaries with a behavioral health visit by year in Texas, 2009 through 2017**

Source: Mathematica's analysis of TAF data from Texas from 2009 through 2017.

DSRIP = delivery system reform incentive payment.

Figure B.24. Percentage of beneficiaries with diabetes who had an HbA1c test by year in Texas, 2009 through 2017

Source: Mathematica's analysis of TAF data from Texas from 2009 through 2017.
DSRIP = delivery system reform incentive payment.

Figure B.25. Percentage of ED visits for ACSCs with a follow-up visit within seven days by quarter in Texas, January 2009 through December 2017
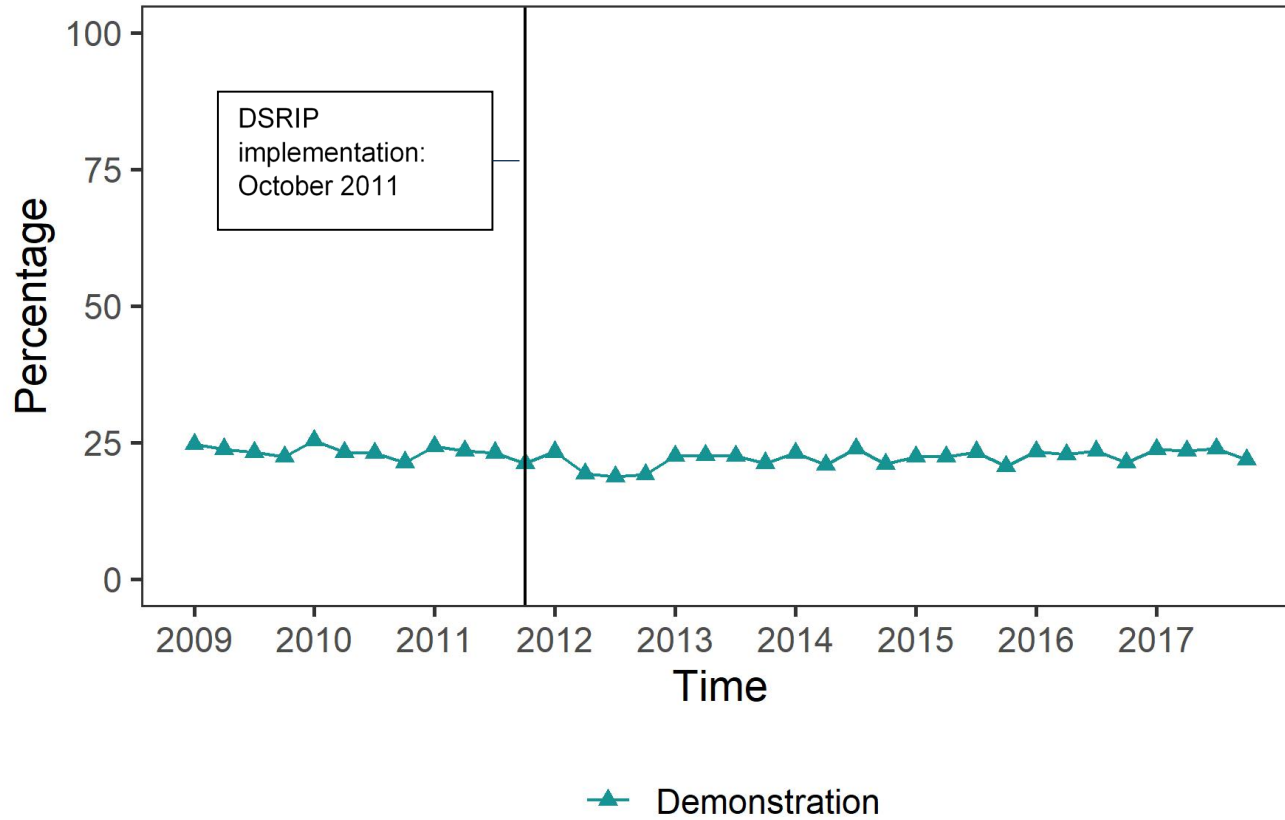
Source: Mathematica's analysis of TAF data from Texas from 2009 through 2017.

Note: We dropped data from 2013 Q4, the quarter before the state transitioned to reporting Medicaid data through T-MSIS.

DSRIP = delivery system reform incentive payment.

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

Appendix C:

Sensitivity Analyses

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

Table C.1. Sensitivity analyses for California: difference-in-differences findings compared with interrupted time series findings

	Difference-in-differences estimates		ITS estimates	
	California (DSRIP)	California (PRIME)	California (DSRIP)	California (PRIME)
	Relative change in post-period—DSRIP group	Relative change in post-period—demonstration group	Relative change in post-period—DSRIP group	Relative change in post-period—demonstration group
ED visits	—	—	—	—
Avoidable ED visits	—	—	—	—
Ambulatory care visits	NS	NS	NS	+
Primary care visits	NS	NS	+	NS
Behavioral health visits	NS	+	NS	NS
HbA1c testing	NS	NS	NS	+
Follow-up within seven days after an ED visit for ACSCs	—	NS	—	+

Source: Mathematica's analysis of TAF data from California, Virginia, and Washington from 2009 through 2017.

Note: This table shows two separate models for each outcome of interest. The difference-in-differences estimates show the main outcome of interest, the relative change in the post-period between the demonstration and comparison group after implementing DSRIP and PRIME. The ITS estimates show the main outcome of interest—whether the trend in the outcome of interest changed significantly after implementing DSRIP and PRIME. The models also controlled for the following characteristics measured in 2011: the percentage of beneficiaries who were children, the percentage of beneficiaries with a behavioral health condition, the percentage of beneficiaries with one or more chronic conditions, the percentage of individuals in the HSA who were white, the percentage of individuals in the HSA covered by Medicaid, the percentage difference of individuals in the HSA covered by Medicaid between 2011 and 2015, the percentage difference of individuals who were uninsured in the HSA between the baseline year and 2015, the number of hospital beds per resident in the HSA, the percentage of individuals in the HSA who lived in a primary care shortage area, and the percentage of individuals in the HSA who lived in a mental health shortage area. For quarterly measures, the models controlled for seasonal effects. Although the results are largely consistent across both models, we saw changes related to ambulatory care visits, primary care visits, behavioral health visits, HbA1c testing, and follow-up after an ED visit for ACSCs.

+ = a favorable outcome; — = an unfavorable outcome; NS = not significant; ACSC = ambulatory care sensitive condition; DSRIP = delivery system reform incentive payment; ED = emergency department; HbA1c = hemoglobin A1c; ITS = interrupted time series; PRIME = Public Hospital Redesign in Medi-Cal.

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

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