June 2017

MEDICAID 1115 DEMONSTRATIONS

Implementing Measurement Requirements in DSRIP Demonstration Programs

Julia Baller, Jessica Heeringa, Carey Appold, Keanan Lane, Ryan Stringer, and James Woerheide

Introduction

Delivery System Reform Incentive Payment (DSRIP) programs, authorized as Medicaid section 1115 demonstrations, provide incentive funding to safety net hospitals and other providers, such as physicians and community clinics, for infrastructure investments, quality improvement, and population health advancement. To receive these incentive payments, providers must implement projects that achieve specified milestones and metrics. The projects, milestones, and metrics are broadly designed to build infrastructure and workforce capacity among safety net providers, transform the way services are delivered, and improve the health of Medicaid beneficiaries and the uninsured. In this brief, we focus on the largest DSRIP demonstration programs approved by January 1, 2016 including California, Massachusetts, New Jersey, New York, and Texas. For California, we examine the state's first DSRIP program and its new DSRIP program, the Public Hospital Redesign and Incentives in Medi-Cal (PRIME) program. For Massachusetts, we include the state's initial Delivery System Transformation Initiatives (DSTI) program and its renewed DSTI program.

Although the DSRIP program characteristics differ in each participating state (varying, for example, in terms of the eligible participating providers, maximum potential funding, project focus, and number of different projects), measurement is an essential component of all programs (for details, see Table A-1 in the Appendix). Measurement and reporting activities typically serve three broad purposes. First, measurement activities are a source of program monitoring data for the Centers for Medicare & Medicaid services (CMS), states, and providers. Program monitoring helps stakeholders recognize and resolve operational barriers, identify opportunities for technical assistance, and ensure the program's integrity and accountability. Second, measurement is the basis of provider performance assessment; participating providers regularly report specific milestones and metrics to determine the extent to which they are achieving stated objectives. Although programs use different terms, we define **milestones** as the activities that providers must complete within a specified time frame to receive the incentive payment. For example, a provider may have to deliver documentation of a completed data collection plan within the first six months of the program. We define **metrics** as the quantitative measures, with predefined numerators and denominators, that providers must either report (pay-for-reporting metrics) or improve on (pay-for-performance metrics) to receive incentive funding—for example, avoidable hospital readmissions. Third, measurement is necessary for evaluation. Each DSRIP program must complete an external evaluation of its demonstration to assess the program's effectiveness on a statewide level, which in part relies on the measurement and reporting activities just described. The external evaluations are

THE MEDICAID CONTEXT

Medicaid is a health insurance program that serves low-income children, adults, individuals with disabilities, and seniors. Medicaid is administered by states and is jointly funded by states and the federal government. Within a framework established by federal statutes, regulations and guidance, states can choose how to design aspects of their Medicaid programs, such as benefit packages and provider reimbursement. Although federal guidelines may impose some uniformity across states, federal law also specifically authorizes experimentation by state Medicaid programs through section 1115 of the Social Security Act. Under section 1115 provisions, states may apply for federal permission to implement and test new approaches to administering Medicaid programs that depart from existing federal rules yet are consistent with the overall goals of the program and are budget neutral to the federal government.

Some states have used section 1115 waiver authority to implement delivery system reform incentive payment (DSRIP) demonstrations. Since the first DSRIP program was approved in 2010, the breadth and specific goals of these demonstrations have evolved, but each aims to advance delivery system transformation among safety net hospitals and other Medicaid providers through infrastructure development, service innovation and redesign, and population health improvements. More recent DSRIP demonstrations have also emphasized increasing provider participation in alternative payment models, which intend to reward improved outcomes over volume.

expected to focus on the impact of the program on the triple aim: improve the experience of care, improve the health of populations, and reduce per-capita costs of health care.

The purpose of this brief is to give the reader an understanding of how measurement fits into the broader DSRIP program, describe the way that states and CMS designed the measurement requirements in each program, explore how states and providers are implementing the measurement requirements, and assess key informants' beliefs about how measurement is influencing change. Our goal in this brief is not to assess the adequacy of states' measurement requirements, but insights from this brief can help inform the design of measurement requirements in new programs or program renewals. Future quantitative analyses will be needed to examine the extent to which program design elements, including the measurement requirements, influence the impact of DSRIP demonstration programs.

This brief includes a conceptual framework, which illustrates the role of measurement in DSRIP; a description of the metrics used in DSRIP programs; a report on the barriers and facilitators that key informants say they experienced in implementing state measurement requirements; a discussion of how measurement has influenced DSRIP achievements; and the key lessons learned and recommendations. We conclude by discussing the implications of these findings for the upcoming impact evaluation of DSRIP programs.

The Role of Measurement in DSRIP

Goals of DSRIP programs. Measurement plays a key role in achieving the broader goals of DSRIP programs (Figure 1). To some extent, each program is designed to achieve system transformation by addressing the following proximal goals:

- Building infrastructure and workforce capacity among providers serving individuals who are covered by Medicaid or are uninsured
- Strengthening collaborations and coordination among providers serving individuals who are covered by Medicaid or are uninsured
- Encouraging service innovation, redesign, and clinical quality improvements among providers serving individuals who are covered by Medicaid or are uninsured
- Promoting strategies to improve population health
- Preparing safety net providers for value-based purchasing and alternative payment models

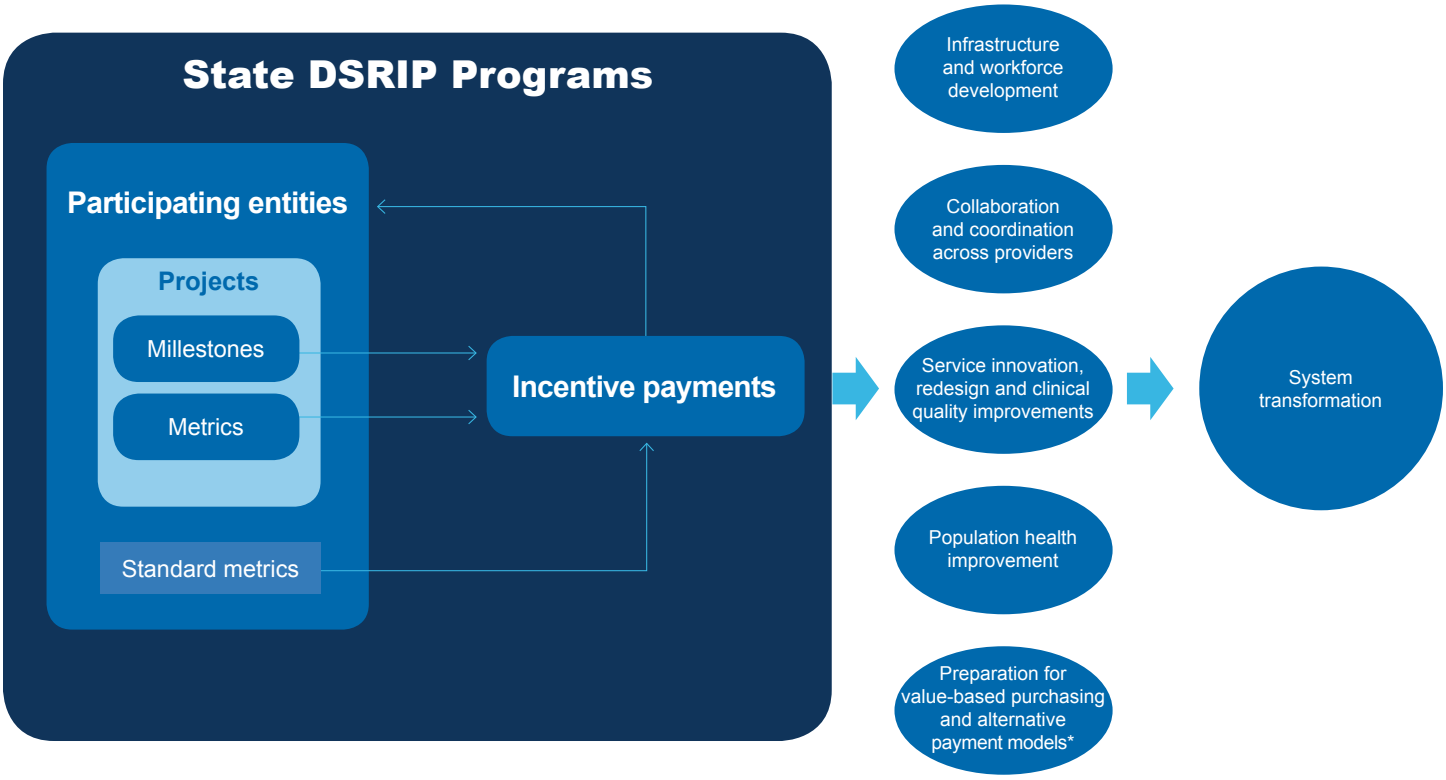
Although system transformation is a clearly stated objective across all programs, the proximal goals are not always explicitly described in the Special Terms and Conditions (STCs) negotiated between CMS and each state. Furthermore, short-term program goals often evolve within a state over the course of its demonstration period and across DSRIP programs over time.

The role of participating entities. To achieve program goals, different entities in each state are responsible for carrying out improvement projects and reporting milestones and metrics. In Massachusetts and California, these entities are hospital systems; in New Jersey, they are hospitals and hospital-led partnerships with community-based providers; in Texas, they are performing providers operating within regional networks; and in New York, they are regional networks.

Using measurement to drive change. The projects carried out by each entity and the milestones and metrics they report are designed to be the vehicles that drive change in DSRIP programs. In all five states, the targeted entities are carrying out projects in three domains: infrastructure and workforce development, service innovation and redesign, and population health.[1] For instance, a common infrastructure project in Texas is the establishment of more primary care clinics; a common service innovation and redesign project in multiple states is integration of primary care and behavioral health services; and a common population health improvement project in New York is the promotion of mental, emotional, and behavioral well-being in communities. The focus of the projects tend to evolve over the demonstration period; early projects tend to focus on infrastructure and workforce development, which provide the framework necessary for projects to eventually focus on service innovation and redesign and population health improvement.

Each participating entity reports milestones and metrics that are explicitly tied to projects. Milestones are a way to identify structural progress and help monitor projects with the goal of supporting future achievement of outcomes. They typically fall within the infrastructure development domain and focus on health information technology, workforce and human resources, and physical infrastructure. Metrics drive monitoring, performance assessment, and evaluation, and they focus on the same domains as the projects.[2] In addition to the milestones and metrics tied to projects, participating entities in California's first DSRIP, Massachusetts' DSTI, New Jersey's DSRIP, and Texas' DSRIP programs report standard metrics that are not tied to projects. For example, all participating hospitals in New Jersey report the same 33 metrics regardless of project selection, including metrics such as comprehensive diabetes care measures and follow-up after hospitalization for mental illness. Together, milestone achievement and metric reporting and performance determine the incentive payments that the participating entity receives.

Figure 1: The role of measurement in achieving DSRIP goals



*California's new DSRIP program (the Public Hospital Redesign and Incentives in Medi-Cal [PRIME]), Massachusetts' DSTI program, and New York's DSRIP program explicitly aim to move toward value-based purchasing and alternative payment models.

DSRIP Metrics

At the program's inception, the state works with CMS to develop a state menu that specifies DSRIP project options, associated milestones and metrics, and standard metrics that are eligible for DSRIP funding. To better understand how states and CMS designed their state menus, we created a database that classifies the quantitative metrics among the states included in this analysis into several categories. A total of 970 quantitative metrics are on these states' menus. Texas has the most metrics on its menu (497), and the Massachusetts DSTI renewal has the fewest (39). In all programs except for New York's (active November 1, 2015 through June 30, 2017), most metrics fall into the service innovation and redesign domain. In California's first DSRIP program, 90 percent of metrics are focused on service innovation and redesign, around 75 percent of metrics fall into this domain in Massachusetts, Texas and New Jersey, and just 45 percent of metrics are in this domain in New York. In contrast, 54 percent of the metrics in New York's menu are focused on population health improvement, and 1 percent focus on infrastructure and workforce development.

Overall, the state menus include more outcomes metrics than process metrics, but this varies within each program. Around 65 percent of the metrics in California, New Jersey, and Massachusetts' DSTI renewal program measure processes, compared with one-third of metrics in Massachusetts' first DSTI, New York's DSRIP, and Texas' DSRIP programs. In contrast, the proportion of metrics that capture outcomes varies from a low of 24 percent in California to a high of 58 percent in New York. To a much lesser degree, state menus also include structural (3 percent), patient experience (8 percent), and cost metrics (2 percent).

Just as the number and focus of metrics vary by state, there is substantial variation from one state to the next in the distribution of metrics by care setting. Metrics focused on ambulatory or outpatient care are the most common measures in all programs, ranging from a high of 61 percent of the metrics in California's DSRIP program to a low of 35 percent in Texas. Hospital inpatient metrics are the second most common, with about one-third of metrics in California, Massachusetts, New Jersey, and Texas focusing on this care setting. In contrast, only four percent of metrics in New York are hospital inpatient metrics. This cross-state variation likely reflects differences in the targeted entities; California, New Jersey, and Massachusetts are hospital-based programs, whereas New York has various types of providers. Surprisingly, inpatient metrics make up the largest proportion of total metrics on a state's menu in Texas even though the Texas DSRIP program is not considered a hospital-

based program. Although hospitals are frequently the lead entity, lead entities in Texas also include community mental health centers, city or county health departments, and physician groups affiliated with academic health centers.

Each state menu includes metrics with a range of credentials. Thirty-five percent of metrics in all programs are endorsed by the National Quality Forum (NQF), an organization that reviews and assesses measures on the basis of their importance, scientific acceptability, usability and relevance, and feasibility to collect ("What NQF Endorsement Means" 2017). Among the 65 percent of metrics that are not NQF-endorsed, five percent had their NQF endorsement removed during the demonstration period, 37 percent are nationally recognized by other well-known measure stewards such as the National Committee for Quality Assurance, and 22 percent are "innovative" metrics that are not recognized by national stewards.

In sum, understanding the design of the state menus is an important first step in investigating implementation of the measurement requirements and ultimately assessing the adequacy of the metrics in supporting system transformation. The quantitative metrics database suggests substantial variation in the types of metrics included in the state menus, which differ in domain focus, type of measure, care setting, and credentials.

Barriers And Facilitators to Implementing the DSRIP Measurement Requirements

Although the projects, milestones, and metrics are designed to drive system change, certain factors facilitate implementation of DSRIP measurement requirements while others act as barriers. These barriers and facilitators (Table 1) may affect whether DSRIP providers can meet the requirements, receive incentive payments, and ultimately achieve program goals. Based on interviews with a wide range of stakeholders, we identified three main areas that can make it easier or harder to implement the measurement requirements: measurement design elements, provider- and state-level factors, and contextual factors.

Table 1. Facilitators and barriers to implementing measurement requirements in DSRIP programs

Factors	California	Texas	Massachusetts	New Jersey	New York
Measurement design elements					
Measurement flexibility	+/-	+/-	-		-
Measurement standardization	+/-	-	+	+	
Alignment with other measurement activities outside DSRIP	+	+/-	+/-	-	-
Project measures centrally calculated by the state		+	-	-	+
Provider- and state-level factors					
Information technology infrastructure	+/-	+/-	+/-	-	+
DSRIP time frame	-	-	-	-	-
Experience with measurement before the DSRIP program	+/-	+	-		+
Performance feedback given to providers	+		+	-	+
Adequate support staff to carry out measurement requirements	-	+/-	+/-	+/-	+/-
Technical advisory committees to help in measure selection	+	+			+
Data accessibility for calculating milestones and metrics		-		-	-
Administrative burden of measurement requirements	-	-		-	-
Learning collaboratives	+	+	+	+	
Contextual factors					
Greater national attention paid to measurement	-	+	-	+/-	
Available evidence-based metrics to include in state menu	-	-		-	

Sources: Mathematica analysis of materials from 15 key informant interviews with representatives from California, Texas, Massachusetts, New Jersey, and New York.

Notes: "+" indicates a respondent from a given state mentioned a characteristics as a facilitator, "-"indicates a respondent from a given state mentioned a characteristic as a barrier or a facilitator that they lacked, and "+-" indicates that some respondents said the factor was a facilitator whereas others indicated it was a barrier. The table includes all barriers and facilitators described by representatives from more than one state. Those barriers and facilitators that were mentioned by a single respondent or multiple respondents from a single state are not included in the table.

Measurement design elements. Respondents across programs underscored a common tension between measurement flexibility and standardization. On the one hand, innovative metrics can be more flexible, intended either to fill in the gaps where endorsed measures are not available or to more accurately reflect activities carried out through individual projects than the metrics prescribed by CMS and the states. However, when providers have considerable flexibility in defining which metrics they report and how they are calculated, which was the case for the first DSRIP programs in California and Massachusetts, the lack of standardization can make it more challenging to evaluate the impact of DSRIP at the state level.

For example, in California's first DSRIP program, designated public hospitals (DPHs) opted most often to implement a project focused on expanding medical homes (Pourat et al. 2014). Providers could choose from several process measures for this project, including the following milestones: increase number of primary care clinics using the medical home model; develop hospital policies on medical homes; and establish a primary care team. In addition, providers chose from several improvement metrics, including, but not limited to: number or percent of eligible patients assigned to a medical home; time to third next-available appointment; and percent of primary care visits at medical home. Although the measurement requirements' flexibility allowed DPHs to tailor their projects and gave them the greatest chance for successfully receiving incentive payments, it posed several challenges to evaluating the program as a whole. By contrast, many Performing Provider Systems (PPSs) in New York are implementing a project to increase the number of primary care practitioners (PCP) with patient-centered medical home (PCMH) certification and/or advanced primary care models. All PPSs implementing this project must report the same 14 metrics, including: potentially avoidable emergency room visits, percent of PCPs meeting PCMH/Advanced Primary Care, the Healthcare Effectiveness Data and Information Set access/availability of care measure, and several Consumer Assessment of Healthcare Providers and Systems measures. The standardization of New York's program lends itself to broader evaluation, but leaves little room for tailoring reporting to specific project activities or a project's target population. Although we expect variation in the level of metric standardization to have implications for system change, it was too early at the time of this analysis to determine the extent to which this is the case.

Many respondents mentioned that the DSRIP measurement requirements posed a substantial administrative burden to both providers and states, and implementation is much easier when DSRIP metrics align with other quality improvement activities. For instance, in Texas, 338 participating providers participate within one of 20 Regional Healthcare Partnerships (RHPs). Each participating provider is carrying out multiple projects, and as of August 2016, a total of 1451 projects are being implemented across the state. The providers must report multiple milestones and metrics for each of the projects, along with a set of standard metrics, twice a year. These metrics are often similar to metrics the providers have to report for other programs, but they have to be calculated using different measure specifications. Aligning the metrics with those collected in other programs not only would improve efficiency, but also ensure consistency in the goals across different programs. Despite the efficiencies that can be generated through measure alignment, this process can introduce ambiguity about how to attribute outcomes and claim associated incentives within and outside the DSRIP program. For example, providers in Massachusetts stated that they have to be careful not to "double-dip" in quality-based payment programs. In addition, using the same measures for different programs poses challenges to attribute outcomes of system transformation and population health to DSRIP.

Provider and state-level factors. Information technology (IT) systems at the provider level were brought up often as a key challenge. Some providers, particularly large academic medical centers, were well versed in measurement and had advanced IT capabilities before they implemented DSRIP, but other provider types were "starting from scratch." For example, many did not have electronic health records (EHRs) in place that enable metric calculation and data analytics, which posed a particular challenge for calculating measures based on information obtained from medical charts. One New Jersey respondent noted substantial differences in the ability to collect data even within a single medical center, which uses a hybrid health records system; some departments and practices use an EHR and others use paper charts. Developing an accessible, efficient system and work flow for both types of records continues to be a challenge for this medical group.

> "The level of sophistication across hospitals varies, like in EHRs, data warehouses, and reporting. Having those basics in place for accurate measurement was a hurdle, especially for smaller hospitals. It has taken time with population-health based measures to report. Despite measure specifications, hospitals will sometimes continue to report measures differently."
>
> —Massachusetts interviewee

Compounding this issue, interview participants overwhelmingly mentioned DSRIP's short time frame as a barrier to their successfully implementing the measurement requirements, particularly for provider systems that initially lacked advanced IT systems. In respondents' views, the limited ramp-up periods of DSRIP programs are not long enough to develop the substantial infrastructure necessary to carry out measurement activities.

Moreover, respondents questioned whether the metric targets, particularly for outcomes metrics, are achievable over the course of such a short demonstration period. Even if the infrastructure is in place to measure system transformation and population health outcomes, it may be unrealistic to assume that three or five years is enough time to move the needle and change the way care is delivered.

> "There was a lack of understanding of how much time would be necessary to implement this in a meaningful way. There were some hospitals that had just one part-time employee. That was humbling. They couldn't collect data and report it. Some had to do manual chart reviews. The labor that went into executing the reporting at both the hospital and state level was huge, and we didn't anticipate it out of the gate."
>
> —California interviewee

At the same time, interviewees remarked that transferring knowledge and communicating performance data through dashboards, establishing web portals, having accessible health information exchanges and regional health information organizations, and participating in learning collaboratives are critical facilitators to measurement efforts. For example, state representatives in New Jersey, New York, and Texas told us that dashboards or web portals that show timely, provider-level performance data make it easier to monitor a program and improve quality, and also give some transparency to the program. But one New Jersey provider contradicted the state representative's claim, sharing the perspective that claims-based metrics (for instance, follow-up after hospitalization for mental illness) that are calculated by the state and included in the web portal are not communicated in a useful way; in their view, the data are not verifiable, not timely, and lack adequate context for the provider to use them to inform future action.

Contextual factors. Participants remarked that the state of the quality measurement field at the national level has been both a barrier and facilitator to implementing DSRIP measurement requirements. States and providers can more efficiently and consistently collect and report performance data when nationally endorsed, well-supported measures are available. New Jersey, for instance, used the NQF and the National Quality Measures Clearinghouse, sponsored by the Agency for Healthcare Research and Quality, to select metrics with strong support systems for use in DSRIP. However, the measurement field has been a moving target, which had significant implications for DSRIP programs. Between 2011, when the first demonstration program began in California, and 2014, when New York's DSRIP program began the field of quality measurement grew substantially. Where California had few nationally endorsed evidence-based measures from which to choose, New York had many more measures to consider. Furthermore, as the measurement field evolved, national organizations such as the NQF often stopped endorsing measures as the evidence base matured (discussed in detail below). States had to decide whether to continue using those measures that providers had grown accustomed to or align the measures with the evidence. Providers found they were ill-equipped to adapt when states opted to drop measures.

> "We had a required project on sepsis and … we needed an intervention everyone knew. However, it turned out that there was no agreed-upon or nationally recognized definition of sepsis or a way to measure it, and every member was doing it slightly differently. It took us a year to develop a consensus with our membership. Then, within a year in DSRIP, we did get a national definition on sepsis and that creates the question of, "'Should we continue what we're doing or adopt the national measure?'"
>
> —California interviewee

How Measurement Has Influenced DSRIP Achievements

Lessons from early-stage reporting. State and provider interviewees in California, Massachusetts, and Texas underscored the importance of early stage milestones in providing a clear, actionable roadmap for achieving the broader goals of DSRIP. Interviewees stressed, however, that the utility of these milestones in achieving system transformation depends on the extent to which they align with a state's DSRIP metrics and projects as well as the broader goals of the program. For example, participants in New Jersey stressed that each participating hospital is only implementing a single DSRIP project, which typically focuses on individuals with a specific chronic disease or condition. The milestones and metrics associated with the projects are well-aligned with the target population, but the standard metrics reported for all hospitals include the broader population of people who are covered by Medicaid or are uninsured. The New Jersey providers voiced their concerns that the fundamental lack of alignment across the project metrics and broader DSRIP goals would hamper their ability to achieve system transformation.

> "In order to move these outcome measures, we need to have smaller, more proximal [milestones] that we can review on a weekly, monthly basis to understand if we're having process improvement."
>
> —Massachusetts interviewee

Furthermore, key informants from these states revealed the critical learning that took place during the early phases of reporting milestones and metrics. States and providers described a steep learning curve when it came to understanding the data and using them to monitor programs and improve quality of care. Respondents in California, Massachusetts and Texas, whose first programs were approved between 2010 and 2012, said they began implementation at a time when the quality measurement field was in flux, far behind where it currently stands. As a result, states and providers spent the first few years building infrastructure and data capabilities while learning the mechanics of calculating metrics and the process for state and federal reporting. This ramp-up period did improve the quality and consistency of the data; however, one provider in Texas stated that these improvements make it difficult to attribute provider progress on the DSRIP metrics to better quality of care and not to higher quality of data. Additionally, providers in California, Massachusetts, and Texas made the point that only now, years into these programs, is measurement beginning to drive changes in the way care is being delivered. As a result, respondents did not point to a single metric or set of metrics that drives quality improvement or system transformation. Given these comments, we expect that initial DSRIP programs in California, Massachusetts, and Texas will have a positive influence on structural metrics, such as EHR use, but minimal influence on DSRIP process and outcomes metrics.

> "DSTI measures have had a ripple effect across the organization in doing careful project management and in understanding and using data, such as the understanding between process and outcomes measures, and understanding accountability for performance."
>
> —Massachusetts interviewee

Perceptions of the link between measurement and change. Respondents suggested that incentivizing provider-level metrics drives providers to focus on DSRIP activities in both positive and negative ways. Overall, respondents expressed the opinion that tying incentive payments to milestones and metrics has made them pay more attention to improving performance on those measures. Although this incentive design may be influencing the way care is delivered, state representatives in Massachusetts and Texas stated that unintended consequences can sometimes arise when efforts are shifted away from activities that are more critical for improving care of the target population but are not explicitly tied to DSRIP funding.

The interviews also shed light on how the design of the measurement requirements is impacted by the evolving goals of DSRIP programs, which likely influences how programs change care delivery. For instance, respondents in California and Massachusetts stressed that their first DSRIP programs implicitly sought to support the safety net and provide financing for the infrastructure development necessary for delivery system reform. As a result, providers were able to select the milestones and metrics to report and set their own performance targets, making it more likely they would reach their targets and receive full incentive payments. In other words, this flexibility reduced the risk of funding loss at the provider level. Providers, in particular, considered this flexibility critical because the payments were fundamental to the capacity building the program set out to achieve. This foundation is now enabling innovations in their subsequent DSRIP programs in the areas of service innovation, population health improvements, and system transformation. In contrast, the DSRIP program goals in New Jersey, New York, and Texas more directly emphasize delivery system transformation. As a result, the metrics in these programs tend to cross more service settings, and there is greater risk for funding loss. The extent to which these elements of the incentive design are influencing service delivery and system transformation, however, is still unknown.

In addition, incentivizing the state-level metrics in the applicable programs (Massachusetts' renewed DSTI and New York's DSRIP programs), was not seen as influencing the activities carried out by providers to date. Instead, performing providers were focused on the metrics that are tied to provider incentive payments, as these have the most immediate impact on their daily operations; incentivized state metrics were rarely a consideration in how providers deliver care. Notably, at the time of this analysis, New York was still in the early stages of implementation, and Massachusetts only introduced incentivized state-level metrics in its renewed DSTI program. As such, it is possible that incentivized state-level metrics will have a greater impact on these programs in the future.

> "At this point in time, it feels largely like the state commitment is in the background of performance measures. We're not hearing dialogue between PPSs on how to best achieve the state's goals. It feels very much that the PPS need to be their own team, and that's how to best help the state."
>
> —New York interviewee

Key Lessons Learned and Recommendations

The findings from the key informant interviews shed light on several aspects of the measurement requirements that CMS and states may consider for the future.

Test new measures. Respondents recognized the importance of standardized metrics for comparing performance of different providers, but they also stressed the need to test innovative measures that more accurately represent activities carried out through the projects. DSRIP demonstrations provide a unique opportunity not only to test new measures, but also to validate nationally endorsed measures in new settings. As is the case in the California PRIME program, the process for testing innovative metrics, including each of the steps required, the entities responsible for each step, and a timeline for moving from testing, to pay-for-reporting, and ultimately to pay-for-performance, should be clearly laid out in the demonstrations' STCs.

Assess provider readiness for system transformation. The variation in infrastructure capabilities (for instance, health IT) between states and providers suggests that there is substantial variation in readiness for system transformation, including readiness to participate in alternative payment models. Currently, this variation is not fully accounted for when setting short- and long-term goals for individual providers. One way to better measure and account for this variation is to include an assessment of the organizational infrastructure when providers conduct needs assessments during the planning phase of the program. Currently, needs assessments carried out under DSRIP typically focus on the health and health services challenges experienced by the target communities. Expanding this activity to incorporate organizational resources and capabilities, including fiscal resources, workforce and human resources, physical infrastructure, health IT (Agency for Healthcare Research and Quality 2012), and penetration of alternative payment models, may help providers make better project choices, establish performance goals, set state priorities for technical assistance needs, and inform CMS expectations about achievable progress on milestones and metrics.

Generate timely, actionable information. Providers also consistently noted that continuous communication of performance data and lessons learned via web portals and learning collaboratives were important facilitators to implementation. Providers noted that this level of engagement improved the quality of the data and made the milestones and metrics more actionable for program monitoring and quality improvement. This feedback suggests that both learning collaboratives and state-run web portals with performance dashboards could be important requirements for future DSRIP programs. Learning collaboratives provide an important avenue for technical assistance at both the state and provider level. Web portals with performance dashboards, particularly those that provide standardized, timely performance feedback on a range of programs, make it easier to monitor provider performance, which can inform practice and program adjustments and allow for cross-state comparisons.

Create alignment within the DSRIP programs. The key informant interviews and the results from analysis of the metrics database underscored the need to align the milestones and metrics with the targeted entities, target populations, project goals, and DSRIP goals to achieve system transformation. Achieving this alignment can be a challenge when the proximal program goals lack clarity and evolve over the course of the demonstration period. Moving forward, as CMS considers the appropriate design of measurement requirements for new DSRIP programs and program renewals, careful attention should be paid to this issue to ensure the overall design of the measure requirements support the achievement of the broader goals of DSRIP programs. Development of a driver diagram, or logic model, that illustrates the inputs, outputs, outcomes, and impacts for each program may be a useful exercise to ensure alignment across the program. As the programs evolve over time, the diagrams will need to be revised to ensure that that the projects, milestones, and metrics are still serving the broader goals of the program.

In addition, several interviewees stressed the importance of aligning the milestones' and metrics' performance targets with the goals of the DSRIP program. In early programs such as those in California and Massachusetts, state representatives in California and providers in both states told us that providing funding to safety net hospitals is an explicit goal of the programs. As such, they expressed skepticism about placing this funding at risk by setting performance targets too high. Understanding how differences in incentive designs influence program outcomes will be the focus of a future issue brief.

Build on initial progress with DSRIP renewals. As CMS considers and negotiates renewals for current DSRIP programs, it is important that the measurement requirements build on the progress demonstrated during the initial demonstration period. Building on initial progress requires a greater focus on metrics as opposed to milestones, and a shift from process to outcomes metrics. In addition, the goals of the projects should be more ambitious, with greater emphasis placed on service innovation and redesign and population health improvement as opposed to infrastructure and workforce development. Further, the performance targets should call for improvements beyond those made during the initial demonstration period.

Consider expanded eligibility requirements. Although the measurement requirements should reflect progress made during the initial period, we expect most program renewals to expand the participant eligibility requirements. The measurement requirements must account for these expansions. For example, the California PRIME program expanded eligibility from solely DPHs to include district municipal public hospitals (DMPHs). The DMPHs did not have the infrastructure building

period or the funding opportunities that were available to DPHs during DSRIP. As a result, the DMPHs may face greater challenges in ramping up their program. To account for this variation in their starting points, the STCs only require DMPHs to carry out one PRIME project, in contrast with the nine projects DPHs must implement. It is too early to determine if reducing the number of projects is an effective way to bring in new entities, but other options might include allowing for different project menus, metrics, or performance targets for newly eligible entities.

Implications for Assessing the Outcomes of DSRIP Demonstrations

To appropriately assess the impact of DSRIP demonstrations, it will be important to examine which program characteristics are affecting change. Measurement requirements are one essential set of program characteristics to consider because they play a key role in motivating provider change and performance improvement. Our analysis of the quantitative metrics highlights substantial variation in the metrics states rely on to monitor, assess, and evaluate the demonstrations. The metrics selected and the relevance of those metrics to a particular program will likely have important implications for how much DSRIP programs can impact infrastructure and workforce development, service innovation and redesign, population health improvement, and ultimately system transformation. Metric variation – and thus the outcomes that were explicitly incentivized – should be accounted for when evaluating these programs to ensure appropriate specification of research questions, selection of comparisons, and interpretation of results in any impact analysis. Further, careful attention must also be paid not only to the initial design, but also to how the requirements are carried out in practice. For example, a program may allow providers carrying out a specific project to select from range of metrics.

Evaluations of DSRIP programs must also account for the context in which each program operates, and how that context influences interpretation of performance on incentivized metrics. For instance, respondents consistently mentioned that it would be much easier to implement DSRIP's measurement requirements if they were aligned with those of other measurement programs, such as accountable care organizations. However, incentivizing measures across programs would make it more difficult to directly attribute changes to DSRIP, particularly if alignment reduced the degree to which metrics tightly reflect outcomes expected in the DSRIP program's logic model. We might expect to find larger changes in outcomes that are directly tied to each DSRIP program and smaller effects for standardized measures that are less relevant to a specific program.

ABOUT THE MEDICAID SECTION 1115 EVALUATION

In 2014, the Center for Medicaid and CHIP Services within the Centers for Medicare & Medicaid Services (CMS) contracted with Mathematica Policy Research, Truven Health Analytics, and the Center for Health Care Strategies to conduct an independent national evaluation of the implementation and outcomes of 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.

The evaluation focuses on four categories of demonstrations: (1) delivery system reform incentive payment (DSRIP) programs, (2) premium assistance, (3) beneficiary engagement and premiums, and (4) managed long-term services and supports (MLTSS). This issue brief is one in a series of short reports based on semiannual tracking and analyses of demonstration implementation and progress. The reports will inform an interim outcomes evaluation in 2017 and a final evaluation report in 2019.

METHODS AND DATA SOURCES

The information in the brief is based on an analysis of (1) the Special Terms and Conditions negotiated between CMS and each state, (2) DSRIP operational protocols, and (3) semi-structured key-informant interviews. Between January and July 2016, the Mathematica team constructed a database of the quantitative metrics included in each program's state menu of projects and metrics, a document that specifies project options, associated milestones and metrics, and standard metrics (the early-stage milestones were not analyzed). Each metric was classified along the following dimensions: metric setting (for instance, ambulatory/outpatient care or inpatient care); measure type (for instance, structure, process, or outcome); and Mathematica domain (infrastructure and workforce development, service innovation and redesign, or population health improvement).

Between March and April 2016, Mathematica staff conducted semi-structured interviews with key informants in the five states that are the focus of this report. Interviewees included state administration officials, state contractors, DSRIP providers, and representatives of provider associations. Contacts for these interviews were identified through state documentation and public source documents.

The team developed interview protocols that included questions on implementation of DSRIP measurement requirements, barriers and facilitators to measurement, and the impact of measurement. The protocols explored metrics that are useful for policymaking, program monitoring, and system transformation. The research team conducted a total of 15 phone interviews with 39 participants in the five states. Two team members read all interview notes to develop an initial framework for organizing and coding the material. Using this framework, the team used a combination of deductive and inductive approaches to develop codes and themes, code all notes line by line, and analyze the data. Writers then supplemented information from these interviews with state and provider documentation.

References

Agency for Healthcare Research and Quality. "2011 National Healthcare Quality and Disparities Report." Washington, DC: Agency for Healthcare Research and Quality, 2011.

Pourat, Nadereh, Gerald F. Kominski, Dylan H. Roby, Ying-Ying Meng, Jack Needleman, Arleen Leibowitz, Erin Salce, Alla Bronshteyn, Max Hadler, Xiao Chen, Nigel Lo, Michael Castro, Katja Nelson, Adriane Wynn, Natasha Purington, Michelle Keller, and Dwight Asuncion. "Interim Evaluation Report on California's Delivery System Reform Incentive Payments Program." Los Angeles, CA: UCLA Center for Health Policy Research, September 2014. Available at http://www.dhcs.ca.gov/provgovpart/Documents/Waiver%20Renewal/AppendixCDSRIP.PDF. Accessed July 2016.

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 Massachusetts Executive Office of Health and Human Services. "Special Terms and Conditions, No. 11-W-00030/1, MassHealth Medicaid Section 1115 Demonstration." Approval Period: December 20, 2011–June 30, 2014, as amended October 1, 2013.

U.S. Centers for Medicare & Medicaid Services and Massachusetts Executive Office of Health and Human Services. "Special Terms and Conditions, No. 11-W-00030/1, MassHealth Medicaid Section 1115 Demonstration." Approval Period: October 20, 2014–June 30, 2019, as amended January 20, 2015.

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

U.S. Centers for Medicare and Medicaid Services and Texas Health and Human Services Commission. "Special Terms and Conditions, No. 11-W-00278/6, Texas Healthcare Transformation and Quality Improvement Program." Approval Period: December 12, 2011–September 30, 2016; amendment approved: February 26, 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 Bridge to Reform Demonstration." Approval Period: November 1, 2010–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." Approval Period: December 30, 2015–December 31, 2020

"What NQF Endorsement Means". Washington, DC: The National Quality Forum, 2017. Available at http://www.qualityforum.org/Measuring_Performance/ABCs/What_NQF_Endorsement_Means.aspx. Accessed June 2016.

Table A.1. DSRIP measurement requirements in participating states

Program	Evaluated entity	Standardized set of metrics not tied to projects	Entities allowed to select performance measures	State framework for DSRIP project implementation	Years of pay-for-reporting metrics	Years of pay-for-performance metrics	Number of standardized metrics reported by each provider
California (DSRIP)	DPHs	Yes	Yes	Category 1: Infrastructure development Category 2: Innovation and redesign Category 3: Population-focused improvement Category 4: Urgent improvement in care Category 5: HIV transition projects	DY2–5	DY2–5	20
California (PRIME)	DPHs, DMPHs	No	Yes	Domain 1: Outpatient delivery system transformation and prevention Domain 2: Targeted high-risk or high-cost populations Domain 3: Resource utilization efficiency	DY1–5	DY2–5	N/A
Texas	Performing providers operating within RHPs	Yes	Yes	Category 1: Infrastructure development Category 2: Program innovation and redesign Category 3: Quality improvements Category 4: Population-focused improvements	DY4–5	DY4–5	37
Massachusetts (DSTI 1)	Participating hospital systems	Yes	Yes	Category 1: Development of a fully integrated delivery system Category 2: Improved health outcomes and quality Category 3: Ability to respond to statewide transformation to value-based purchasing and to accept alternatives to FFS payments Category 4: Population-focused improvements	DY2–3	DY2–3	12
Massachusetts (DSTI 2)	Participating hospital systems; state	Yes	Yes	Category 1: Development of a fully integrated delivery system Category 2: Improved health outcomes and quality Category 3: Ability to respond to statewide transformation to value-based purchasing and to accept alternatives to FFS payments Category 4: Population-focused improvements	DY1–3	DY2–3	9
New Jersey	Participating hospital systems	Yes	No	Stage 1: Infrastructure development Stage 2: Chronic medical condition redesign and management Stage 3: Quality improvements Stage 4: Population-focused improvements	DY2–5	DY4–5	45
New York	PPSs; state	No	No	Domain 1: Overall project progress Domain 2: System transformation projects Domain 3: Clinical improvement projects Domain 4: Population-wide projects	DY2–5	DY2–5	30

Source: Mathematica analysis of:

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 Massachusetts Executive Office of Health and Human Services. "Special Terms and Conditions, No. 11-W-00030/1, MassHealth Medicaid Section 1115 Demonstration." Approval period: December 20, 2011–June 30, 2014, as amended October 1, 2013.

U.S. Centers for Medicare and Medicaid Services. "Massachusetts Section 1115 Demonstration Fact Sheet." Centers for Medicare and Medicaid Services, October 30, 2014. Available at: http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/ma/ma-masshealth-fs.pdf.

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

U.S. Centers for Medicare and Medicaid Services and Texas Health and Human Services Commission. "Special Terms and Conditions, No. 11-W-00278/6, Texas Healthcare Transformation and Quality Improvement Program." Approval period: December 12, 2011–September 30, 2016; amendment approved: February 26, 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 Bridge to Reform Demonstration." Approval period: November 1, 2010–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." Approval period: December 31, 2015–December 31, 2020.

Notes: DSRIP = Delivery System Reform Incentive Payments; DPH = Designated Public Hospitals; DY = Demonstration year; PRIME = Public Hospital Redesign and Incentives in Medi-Cal; DMPH = District and Municipal Public Hospitals; RHP = Regional Health Partnership; DSTI = Delivery System Transformation Initiatives; PPS = Performing Provider System.

Table A.2. DSRIP programs' metrics: examples

	California[3]	Texas	Massachusetts	New Jersey	New York
Mathematica domains					
Infrastructure and workforce development	-	Number of practicing psychiatrists per 1,000 individuals	Average panel size for the practice	Children and adolescents' access to primary care practitioners	Percent of primary care practices meeting primary care medical home recognition
Service innovation and redesign	Child weight screening	Diabetes care: foot exam	ED wait time	Follow-up after hospitalization for mental illness	Asthma medication ratio
Population health improvement	Uncontrolled diabetes	Cavities: adults	Low birth weight rate	CD4 T-Cell count	Maternal mortality rate per 100,00 births
Type of measure					
Cost	-	Per episode cost of care	Report of claims-based utilization data for targeted populations	-	Medicaid spending on ED and inpatient services
Outcome	Pediatric body mass index	Patient fall rate	Readmissions	Controlling high blood pressure	Potentially avoidable emergency room visits
Patient experience	Shared decision making	General satisfaction	Percentage of patients who reported their nurses "always" communicated well	-	H-CAHPS: Care transition metrics
Process	Influenza immunization	Well-child visits in the first 15 months of life	Percent of infants delivered vaginally with shoulder dystocia	Ischemic vascular disease: use of aspirin or another antithrombotic	Initiation and engagement of alcohol and other drug dependence treatment
Structural	-	Third next available appointment	Urgent care volume	-	Percent of eligible providers with participating agreements with RHIOs
Care setting					
Ambulatory/outpatient	Child weight screening	Antidepressant medication management	Depression utilization of the PHQ-9 tool	Weight assessment and counseling for nutrition and physical activity for children/ adolescents	Adherence to antipsychotic medications for people with schizophrenia
Emergency department	-	ED visits in which patients left without being seen	Frequent user ED visits	Ambulatory care—ED visits	ED use by uninsured
Hospital inpatient	Sepsis mortality	Risk adjusted diabetes short term complication admission rate	Falls per thousand patient days	CLABSI event	PDI # 14 Pediatric Asthma
Population/community	-	Adolescent tobacco use	Breast cancer screening	Children ages 6–17 years who engage in weekly physical activity	Maternal mortality rate per 100,000 births
Post-acute, long-term care	-	Hospice and palliative care—pain assessment	CMS skilled nursing facility days for target population	-	Percent of long stay residents who have depressive symptoms
Metric credentials					
NQF-endorsed	Thrombolytic therapy	Controlling high blood pressure	Hypertension plan of care	Breast cancer screening	Prenatal and postpartum care—timeliness and postpartum visits
Nationally recognized	Diabetes, short-term complications	Annual monitoring for patients on persistent medications- diuretic	ED wait time: door to diagnostic evaluation by qualified medical personnel	Mental health care utilization	Cholesterol management for patients with cardiovascular conditions

(continued)

	California[3]	Texas	Massachusetts	New Jersey	New York
Endorsement removed during waiver	Medical visits	Appropriate testing for children with pharyngitis	Low birth weight rate: number of low birth weight infants per 100 births	Bipolar disorder and major depression: appraisal for alcohol or chemical substance use	Comprehensive diabetes care—LDL control (< 100mg/dL)
Other, "innovative"	Percent compliance with elements of the Sepsis	Decrease in no. of mental health admissions and readmissions to criminal justice settings such as jails or prisons	Percent of aligned patient population reimbursed under global payment arrangement	-	Medicaid spending on PC and community based behavioral health care

Source: Mathematica DSRIP Metrics Database, generated from a review of:

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 Massachusetts Executive Office of Health and Human Services. "Special Terms and Conditions, No. 11-W-00030/1, MassHealth Medicaid Section 1115 Demonstration." Approval period: December 20, 2011–June 30, 2014, as amended October 1, 2013.

U.S. Centers for Medicare and Medicaid Services. "Massachusetts Section 1115 Demonstration Fact Sheet." Centers for Medicare and Medicaid Services, October 30, 2014. Available at: http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/ma/ma-masshealth-fs.pdf.

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

U.S. Centers for Medicare and Medicaid Services and Texas Health and Human Services Commission. "Special Terms and Conditions, No. 11-W-00278/6, Texas Healthcare Transformation and Quality Improvement Program." Approval period: December 12, 2011–September 30, 2016; amendment approved: February 26, 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 Bridge to Reform Demonstration." Approval period: November 1, 2010–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." Approval period: December 31, 2015–December 31, 2020.

Notes: ED = emergency department; HCAHPS = Hospital Consumer Assessment of Healthcare Providers and Systems; RHIOs = Regional Health Information Organization; PHQ = Patient Health Quationnaire; CLABSI = Central line-associated bloodstream infection; PDI = Pediatric Quality Indicators; CMS = Centers for Medicare and Medicaid Services; NQF = National Quality Forum; LDL = low-density lipoprotein; PC = primary care; indicates that the state's menu does not include a metric in the respective category.

Endnotes

[1] In New Jersey, hospitals carry out one project that goes through three stages: (1) infrastructure development, (2) service innovation and redesign, and (3) population health improvement.

[2] Although the projects and metrics focus on the same domains, the two components of the project/ metrics dyad are not always in the same domain. For example, participating entities may have to report metrics that fall within the service innovation and redesign domain for a project focused on population health improvement.

[3] California's first DSRIP program included many milestones focused on infrastructure development. However, they were not quantitative metrics, and as a result, were not included in our analysis of metrics.