



Determining the Impact of State Demonstrations: Considerations for State and Federal Policymakers

Amy Clary



This state health policy briefing captures key themes that emerged during a June 2015 meeting of state and federal leaders convened by the National Academy for State Health Policy (NASHP) with the support of The Commonwealth Fund. The goal of the meeting was to identify actionable steps toward a shared federal and state agenda on improving the monitoring and evaluation of state demonstrations, with the goal of capitalizing on successes while easing administrative burden and providing useful data for policymakers.

Introduction

States and the federal government are investing heavily in state demonstrations to reform the health care delivery system.ⁱ As part of national health care reform, the federal government is charged with testing innovations in the health care system that save money while maintaining or improving the quality of health care.ⁱⁱ Many of these innovative models are tested through demonstrations, which require investment from the state. State and federal policymakers need evaluation data to justify these investments to legislators and other stakeholders, and to accurately assess the effectiveness of the demonstrations.

Designing state demonstrations that meet the needs of both state and federal policymakers—as well as consumers—can be challenging. For their part, states seek the flexibility to tailor demonstra-

tion programs to their specific needs. However, too much flexibility in program design can jeopardize the standardization federal officials need to assess the impact of demonstrations on a national level.

In overseeing and assessing these demonstrations, both federal and state policymakers have two overarching goals: to conduct evaluation and to carry out monitoring. Evaluation is necessary to measure the quality and cost impacts of demonstrations, to identify mid-course improvements to interventions, and to identify and spread the most promising practices that emerge from demonstration projects. Monitoring is necessary to assure on a relatively real-time basis that beneficiaries are receiving services consistent with program intent and that public dollars are being properly spent, and to assess directionally how implementation is going. While data requirements for impact analyses require a longer trajectory, they, too, are needed to inform policy, as well as protect beneficiaries and public dollars. Greater alignment across programs

Distinguishing Evaluation From Program Monitoring

Evaluation and program monitoring are methods used to measure the impact of a demonstration or grant program. The U. S. Government Accountability Office (GAO) defines program evaluations as “systematic studies that use research methods to address specific questions about program performance” (U. S. GAO, June 2013). Program monitoring generally uses self-reported measures to track progress toward goals. It informs program management and drives changes in implementation when necessary to meet program goals.

and between state and federal evaluation and monitoring activities could streamline reporting requirements and improve the effectiveness and efficiency of monitoring and evaluation.

A number of promising strategies and areas of agreement emerged from the June 2015 discussion between state and federal officials.

- A **planning period** at the start of demonstrations could help states finalize implementation protocols and prepare for ramp-up, and could help states and federal agencies align performance measures and evaluation criteria, and assure that they are consistent with the goals of relevant programs.
- A meaningful, routinely reported, jointly developed set of **common core performance measures** could help states and federal officials streamline reporting and evaluation.
- **Communication** between state and federal officials early in the development and design of potential demonstrations could help align state and federal program and evaluation goals.
- **Including independent evaluation experts** in the demonstration planning process could help assure that reporting requirements are well aligned with both the goals of the program and the need for rigorous and meaningful evaluation results that policymakers can use.
- **Shorter, more focused monitoring and assessment reports** with key findings in the executive summary would be useful to stakeholders.

State and Federal Officials Use Program Reporting and Evaluation to Improve Programs and Justify Investment to Legislators and Partners

States and the federal government invest substantial time and resources into testing health care reforms. According to the Congressional Research Service, more than half of all federal grant money to state and local governments in 2015 was for health care.ⁱⁱⁱ Strong program assessments can help ensure those dollars maximize desired outcomes. In fact, the White House emphasized evaluation as a way to target funds to programs that work.^{iv}

States and the federal government must justify investments in state demonstrations to legislatures and partners. State

and federal officials alike are accountable to a number of stakeholders to show that public funds used to test health reforms are well spent.

- The Centers for Medicare and Medicaid Innovation (CMMI), the Center for Medicaid and CHIP Services (CMCS), and the other federal agencies that oversee demonstrations are accountable to Congress, the Office of Management and Budget (OMB), and the public to show the value of those demonstrations. Moreover, CMMI operates under a statutory mandate to evaluate models and must provide definitive answers on the effectiveness of the tested interventions. If a model is shown to be effective, the Secretary of Health and Human Services has the authority to expand it. Medicaid demonstrations under 1115 waiver authority are also required to be evaluated by states.
- State officials are accountable to their legislatures, to the federal government, and to the providers, beneficiaries, and other partners who are substantially invested in the success of demonstrations.

State investment is amplified by states' participation in multiple demonstrations that are sometimes overseen by different federal agencies. This contributes to reporting burden for states, which can be exacerbated by redundant requirements for different programs.

- While the abundance of demonstration activity makes it challenging for states and their federal partners to isolate the impact of any one demonstration, it is also a tribute to the energy and commitment to reform evidenced by states and the federal government.

Evaluation and program monitoring help gauge the impact of demonstrations, but often with different aims and methods. Aligning data sources for evaluation and monitoring, when possible, would help ease the burden on states. However, monitoring and evaluation often serve different purposes and seek to answer different questions.

- Formal evaluations are independent assessments designed to evaluate program success and failure, primarily using outcomes measures. They look at big questions that often take time to address. The U. S. Government Accountability Office (GAO) defines program evaluations as “systematic studies that use research methods to address specific questions about program performance.”^{v,vi}

“Administrative and reporting burden leaves less time for collaboration and coordination between programs.”

– State Evaluation Team Lead

- In contrast, program monitoring generally uses self-reported process measures to track progress toward goals, although it can also include process and utilization measures generated from claims data. Monitoring data is not typically compared to the performance of a control group, but might instead be compared to an external benchmark. Monitoring informs program management and drives changes in implementation when necessary to meet program goals. The GAO says performance measurement “entails the ongoing monitoring and reporting of program progress toward pre-established goals.”^{vii}

State demonstration leaders tend to be more burdened by the frequent reporting associated with program monitoring, especially when monitoring is not aligned with evaluation. Cross-state compacts to work on demonstrations and evaluations may help streamline monitoring and evaluation across states. According to one federal official, this was a successful strategy in the Children’s Health Insurance Program Reauthorization Act of 2009 (CHIP-RA).

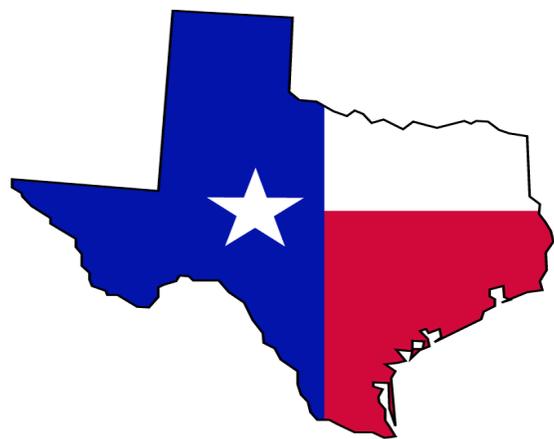
Coordination and Alignment Between and Among State and Federal Partners Can Minimize Reporting Burden and Improve Programs

State demonstrations involve the participation of many different entities, from the federal and state agencies administering and implementing the program, to the governing bodies allocating the funds, to the providers and beneficiaries participating in it. When these entities remain siloed through ineffective communication, demonstrations can suffer. By contrast, programs can benefit when key players work together to determine the goals of the demonstration and the methods through which those goals can be measured.

State and federal officials can work cooperatively to align program goals and reporting and evaluation requirements. Communication between state and federal officials early in the demonstration process may help align state and federal program and evaluation goals, and help balance states’ desire for flexibility with the need for accountability.

State Snapshot: Texas Case Study

Texas collects and reports many different quality measures for the multiple demonstrations in which it participates. We examined six demonstrations in the state, and found little common reporting across programs. Of the quality measures examined, most were reported for one, two, or three of the six demonstrations. In addition to quality measures, the Texas Delivery System Reform Incentive Payment (DSRIP) and Duals demonstrations also used pay for performance measures, including over 300 approved measures for at least 1400 approved DSRIP projects in Texas. It should be noted that Texas had significantly more projects than any other DSRIP program. This case study helps illustrate the scope and complexity of information that states may be asked collect and report.



See the case study documents on the NASHP website:

- [Table of Selected Evaluation Quality Measures and Data Sources for Six Texas Demonstrations](#)
- [Overview of Six Texas Demonstrations](#)

- Federal officials suggest that states approach them with a reform goal, and then work collaboratively to jointly identify the most appropriate funding mechanism. For their part, states seek the flexibility to address state-specific needs in their demonstrations.
- State officials need to know whether they can use their existing resources to replicate successful demonstrations from other states. One state official said, “I need to know not only if a demonstration worked, but if we can actually do it. Is the demonstration part of day-to-day work, or is it an additional layer?”
- It would also give evaluations room to adapt to program changes that occur upon start-up. Without a planning period, evaluations rooted in a program’s baseline concept may lack the nimbleness to adapt to program changes that occur early in the planning or implementation process.

A planning period at the start of demonstrations would help states finalize implementation protocols and prepare for ramp-up, and better align performance measures and evaluation criteria with the goals of the program. Demonstrations are often underway in states before an outside evaluator is hired, which stymies efforts to embed evaluation into the design of the demonstration.

- A planning period would support greater use of driver diagrams and logic models to better identify the elements that demonstrations seek to change and the interventions necessary to change them.
- **Augmented staff capacity and infrastructure would help states and federal partners manage their reporting and evaluation burden.** State and federal officials often struggle to meet the burgeoning demands of monitoring and evaluating demonstrations. Many of the same staff involved in monitoring and evaluation also implement programs and write or respond to new program proposals. Without the appropriate federal capacity to absorb and respond to state monitoring reports, state efforts to collect and report data may be for naught.
- Enhanced IT infrastructure may help existing state and federal staff more efficiently manage the volume of data involved in monitoring and evaluation activities.
- One state official suggested that accompanying CMS data reporting requirements with an enhanced match—beyond the 50 percent federal match to cover the administrative costs of evaluation—could help states better collect

Practical Strategies for Reducing Burden Through Inter- and Intra- Agency Coordination

- **When using the same contractor to evaluate multiple state demonstrations, the federal government can require the contractor to review its work plans in states with an eye toward minimizing burden.**
- **One state was involved in a demonstration sponsored by two agencies in the federal Department of Health and Human Services. When the state made one of the agencies aware of its involvement in the other demonstration, the two federal agencies worked together to coordinate their site visits to the state. These joint site visits minimized the burden to the state.**
- **One federal agency hosts a series of collaboration meetings for their evaluation staff leads to talk about evaluation activities in states and identify areas of overlap.**
- **One federal entity embeds an evaluation staffer in every model team before a request for applications is issued.**



State Snapshot: Vermont Blueprint for Health Profiles

Vermont Blueprint for Health generates profiles for all health service areas in the state comparing their performance on a set of core measures. These profiles use measures from the CMS core ACO measures set, as well as data from the state's all-payer claims database, the DocSite clinical database, and an annual Department of Health behavioral risk factor survey. The state plans to incorporate additional data sources in the future. By using routinely reported data, the profiles drive learning and quality improvement at the provider and system levels without imposing a large reporting burden.

and report claims, encounter, and other data.

Measuring innovation—not simply activity—is challenging. Working cooperatively early in the planning process can help states and federal officials design assessments that pinpoint activities that catalyze change. Currently, federal reporting requirements do not always capture the behavior the demonstration is meant to affect, making it more difficult to determine what works in reform efforts.

Policymakers Need Timely, High-Quality, and Meaningful Data to Measure Innovation and Drive Program Improvement

State and federal policymakers rely on data to help them determine which interventions are driving change and innovation, and which are not. Data systems that include standardized core measure sets—and that routinize the collection of those data—could help ease reporting burden, produce timely evaluations, and support leveraging cross-state learning to help programs improve.

Jointly developing a meaningful set of standardized core measures could help state and federal officials streamline monitoring and evaluation. State and federal officials could work together to jointly develop a modest set of core measures that are common across demonstrations, with additional demonstration-specific measures reported separately as needed. The Congressionally mandated set of adult and child Medicaid core measures could be used as a starting point. For both the minimum core set and the demonstration-specific set, it should be a priority to select from standardized metrics. A set of common measures would relieve some of the burden associated with collecting and reporting different measures for different programs.

- Using routinely reported core performance measures to inform monitoring and evaluation activities would diminish state burden and may also help to better align both evaluation and performance measurement with the goals of the demonstration.
- Maintaining core measures that are consistent across states would also facilitate cross-state comparisons. One state official reported that some states currently modify existing core set measures when reporting on their demonstrations, which makes cross-state comparisons difficult.

State and federal policymakers are considering ways to improve the utility of monitoring and evaluation data. State and federal policymakers acknowledge that shorter quarterly reports highlighting key measures are often more useful than lengthier annual reports for identifying trends and driving timely changes in operations.

- Highlighting key information in an executive summary or through the use of graphs or other visual techniques can also make the data more useful for program implementation.

Using existing standardized data sources for monitoring and evaluation may also produce more timely and usable data.

- One federal official advocated using existing standardized core data infrastructure for multiple purposes, such as admission-discharge-transfer data feeds, and clinical health information exchange data. Meaningful use data, clinical quality measures, and data reported for SIM and other demonstrations could also be incorporated into reporting and evaluation.
- There is hope that the Transformed Medicaid Statistical Information System (T-MSIS) will improve the collection and standardization of data that could be used in monitoring and evaluation. T-MSIS is meant to enhance the ways in which states submit claims, encounter, and other data; and to serve as the foundation of a national data infrastructure.^{viii}

States need timely feedback on their demonstrations.

When evaluation reports are issued after a lengthy time lag—sometimes after demonstrations are over—states cannot use the information to help improve programs. One state official said, “What’s in the evaluation is helpful, but the timing of when it comes out is tough.” Timing evaluations to drive learning and change would increase the impact of evaluations on state demonstrations.

- The federal emphasis on rapid cycle assessment helps produce more timely information about state demonstrations. Although rapid-cycle assessments run the risk of drawing erroneous conclusions from too few data points, their timeliness has great potential to states seeking to course-correct during implementation to help improve program development in response to timely data.

“If data were aggregated routinely and core measure results were produced routinely, evaluation would be a routine production and not an add-on. That’s where capacity investments need to be.”

– State Official

Looking Ahead

Several strategies for improvement emerged from the June 2015 meeting.

State and federal partners could work together earlier in the demonstration process to match state goals with federal funding vehicles, and balance state needs with the accountability needed to drive meaningful change in delivery systems.

- A planning period would help states design evaluations that align with the goals of the demonstration.
- A jointly developed core set of routinely reported meaningful measures that is common across states and programs could help state and federal officials streamline reporting and evaluation.
- Using an agreed-upon set of standardized core measures could ease states' reporting burden and ensure that the information that states report is used, and can be leveraged from other states, to inform decision-making and drive change.
- States seek flexibility in designing demonstration programs that meet their particular needs.
- Policymakers need data systems that provide the timely, complete information needed to gauge the impact of state demonstrations.

“Short term, year-to year political and budget cycles are the biggest challenges to staying the course for long-term change.”

– State Official

Conclusion

Opportunities exist for states and the federal government to work collaboratively to streamline monitoring, reporting, rapid cycle assessment, and longer-term evaluation. In the wake of significant changes to the U. S. health care system, policymakers are closely watching the outcomes of federal demonstrations to see what works and what doesn't in the realm of payment and delivery reform. The stakes are high, and state and federal officials can work together to meet the challenge.

Endnotes

- i For example, the ACA appropriated \$10 billion dollars for the Center for Medicare and Medicaid Innovation (CMMI), which funds innovative payment and delivery models in states and health care facilities. According to a March 6, 2015 Congressional Research Service report, CMMI was appropriated \$10 billion for FY2011-FY2019 and \$10 billion for each subsequent ten-year period. (U. S. Congressional Research Service, Appropriations and Fund Transfers in the Affordable Care Act (ACA), March 6, 2015. Available at: <http://www.fas.org/sgp/crs/misc/R41301.pdf>)
- ii CMMI is tasked with testing innovative health care delivery and payment models: <http://innovation.cms.gov/About/index.html>
- iii U. S. Congressional Research Service, Federal Grants to State and Local Governments: A Historical Perspective on Contemporary Issues, March 5, 2015. Available at <http://fas.org/sgp/crs/misc/R40638.pdf>.
- iv U. S. Office of Management and Budget, Increased Emphasis on Program Evaluations, Memorandum for the Heads of Executive Departments and Agencies, October 7, 2009. Available at https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-01.pdf.
- v Similarly, the United States Code defines program evaluation as “an assessment, through objective measurement and systematic analysis, of the manner and extent to which Federal programs achieve intended objectives” [31 U.S.C. §1115(h)(12)].
- vi U.S. Government Accountability Office, Program Evaluation: Strategies to Facilitate Agencies’ Use of Evaluation in Program Management and Policy Making (GAO- 13-570) (Washington, DC: U.S. Government Accountability Office, 2013), p. 4. Available at: <http://www.gao.gov/assets/660/655518.pdf>.
- vii Ibid.
- viii See the August 23, 2013 CMS letter to state Medicaid directors: <http://www.medicaid.gov/Federal-Policy-Guidance/Downloads/SMD-13-004.pdf>

About the National Academy for State Health Policy:

The National Academy for State Health Policy (NASHP) is an independent academy of state health policymakers working together to identify emerging issues, develop policy solutions, and improve state health policy and practice. As a non-profit, non-partisan organization dedicated to helping states achieve excellence in health policy and practice, NASHP provides a forum on critical health issues across branches and agencies of state government. NASHP resources are available at: www.nashp.org.

Acknowledgments:

The author wishes to thank the advisory committee that helped plan this meeting, the participants at the June 2015 meeting, and the reviewers of this brief. We were fortunate to have the following states, federal officials, and private organizations represented at the meeting: Colorado, Idaho, Iowa, Rhode Island, Texas, Vermont, the Agency for Healthcare Research and Quality, the Center for Medicaid & CHIP Services, the Center for Medicare & Medicaid Innovation, the Centers for Medicare & Medicaid Services, the Office of Management and Budget, and the National Association of Medicaid Directors. The author also thanks The Commonwealth Fund—particularly Eric Schneider, Senior Vice President for Policy and Research; and Rachel Nuzum, Vice President for Federal and State Health Policy—for supporting this project. Any errors or omissions are the author’s.