

An Engine of Change? The Affordable Care Act and the Shifting Politics of Demonstration Projects



PHILIP ROCCO  AND ANDREW S. KELLY

The Patient Protection and Affordable Care Act (ACA) made a number of institutional investments intended to expand the federal government's ability to scale successful demonstration projects into national policy. This article considers how the ACA has reshaped the politics of programmatic innovation in Medicare and Medicaid. A qualitative synthesis of demonstration results suggests that, although the ACA has removed several important veto points to the expansion of successful demonstration projects, numerous barriers to the scaling of reforms remain. These barriers include procedures and techniques that make it difficult to certify the "success" of demonstrations yet make their limitations highly legible. From the analysis, this article draws several lessons for future efforts at delivery-system and payment reform, as well as understandings of policy learning and innovation.

Keywords: Medicare, Medicaid, Affordable Care Act, demonstration projects, policy learning, payment reform

The fragmented character of the American state often acts as a barrier to health policy innovation. Institutionalized veto points raise the transaction costs of policy change through legislative "big bangs." Yet even when policymakers take a more incremental approach to innovation, the path is steep. Throughout the 1990s and early 2000s, federal agencies undertook numerous demonstration projects aimed at controlling costs and improving quality within Medicare and Medicaid. Although these projects yielded promising results, translating them into national policy required congressio-

nal approval. As Andrew Kelly and Philip Rocco (2019) note, given intense opposition from organized medicine and a lack of institutional capacity, promising reforms that aimed to control costs and improve the quality of care have been routinely shelved.

The Patient Protection and Affordable Care Act (ACA), however, has made a number of institutional investments intended to expand government's ability to scale successful demonstration projects into national policy (Guterman et al. 2010). Most important, it consolidated demonstration authority within a new

Philip Rocco is assistant professor in the Department of Political Science at Marquette University. **Andrew S. Kelly** is assistant professor in the Department of Health Sciences at California State University, East Bay.

© 2020 Russell Sage Foundation. Rocco, Philip, and Andrew S. Kelly. 2020. "An Engine of Change? The Affordable Care Act and the Shifting Politics of Demonstration Projects." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 6(2): 67–84. DOI: 10.7758/RSF.2020.6.2.03. Direct correspondence to: Philip Rocco at philip.rocco@marquette.edu, Department of Political Science, Marquette University, P.O. Box 1881, Milwaukee, WI 53201.

Open Access Policy: *RSF: The Russell Sage Foundation Journal of the Social Sciences* is an open access journal. This article is published under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Center for Medicare and Medicaid Innovation (CMMI). Not only did the ACA provide an expansive budget for testing new payment and delivery models, it gave the U.S. Department of Health and Human Services (HHS) the authority to expand demonstrations throughout the Medicare and Medicaid programs without seeking congressional approval—as long as they reduced spending (without reducing quality of care) or improved quality (without increasing spending).

From the beginning, CMMI appeared to create the potential for rapid development and scaling of demonstration models. Congressional committee reports suggested that the new rules would allow for “nationwide implementation” of successful demonstrations (U.S. Congress 2009, 662). The expansion of these models was also projected to result in billions of dollars of savings (CBO 2009). Observers soon identified CMMI as a vehicle for rapidly introducing innovation into the health system (Shrank 2013). Simultaneously, critics of CMMI worried that it would allow for the testing of models that radically restructured both Medicare and Medicaid (Appleby 2016).

Yet although CMMI appeared to give HHS expansive new authority to implement nationwide changes, only two new models have thus far been actuarially certified for expansion—the Pioneer Accountable Care Organization (ACO) Model and the Medicare Diabetes Prevention Program. Only the latter has been formally expanded. Thus the Congressional Budget Office (CBO) has continued to project that CMMI demonstrations will net \$34 billion in savings by 2026, but lacks pertinent information on parameters that may affect those projections, including evidence about the standards CMS’s chief actuary will use in certifying demonstrations for expansion (Hadley 2016).

These results raise an important question: to what extent has the creation of CMMI and attendant expansion of demonstration authority facilitated the implementation of new payment and delivery models within Medicare and Medicaid? What barriers to innovation remain? To investigate the ACA’s effects on the politics of programmatic innovation, this article draws on a qualitative analysis of evaluation reports that maps the outcomes of dem-

onstration models carried out by CMMI and completed between 2012 and 2018 ($n = 14$). Our prior work focused mainly on the factors that affect the scaling of demonstration programs in the pre-ACA period (Kelly and Rocco 2019). Building on that work, this article considers factors that contribute to and inhibit the expansion of Medicare and Medicaid demonstrations under the ACA (Kelly and Rocco 2019; Rocco, Kelly, and Keller 2018). Consistent with Helen Levy, Andrew Ying, and Nicholas Bagley’s review of the ACA (2020), we find that CMMI has endured and is more or less operating as envisioned in the law. Yet we focus our attention on whether CMMI can accomplish the ambitious goals set out for it. The analysis here suggests that greater bureaucratic discretion and fiscal authority have been insufficient conditions for the development of autonomous programmatic innovation (Carpenter 2001). In contrast to the work of Lisa Beauregard and Edward Miller (2020) on state Medicaid waivers, we show that partisan ideology and weak institutional capacity are not the only barriers to policy innovation. Instead, our analysis shows that the ACA’s actuarial certification requirements—which require the CMS actuary to provide a probing analysis of demonstrations—may place new limits on the scope of innovations that CMMI undertakes and is able to expand. Actuarial certification, therefore, acts as a new kind of barrier in the path to innovation and programmatic reform. Other barriers that existed prior to CMMI and the ACA, including limited time horizons, difficulty isolating treatment effects, and limited capacity to address infrastructural gaps, also remain—demonstrating that CMMI’s institutional design not only erected a new barrier to innovation in the form of actuarial certification, but also did not adequately address obstacles that existed prior to the ACA. Models that involve complex interactions between providers, facilities, states, and the federal government present challenges for the attribution of savings and the measurement of quality. Thus, although these models show positive outcomes for care quality and system transformation, their benefits are not easy to calculate under the ACA’s rules. Also, even under CMMI’s authority and subject to new rules, potentially “effective” innovations are not expanded. Still, de-

spite the continued challenges and barriers to scaling demonstration projects, CMMI has greater capacity for iterative learning and learning in the absence of certification.

DEMONSTRATION PROJECTS AND THE POLITICS OF TRIAL AND ERROR

Demonstration projects are an emblematic feature of contemporary American governance. They have their origins in the Progressive Era idea that public policy should be flexible to changing demands for performance, open to bargaining and negotiation, and guided by technical expertise (Orren and Skowronek 2017). Demonstrations flourished during the New Deal and the Great Society as a way of testing novel policy ideas in the absence of large-scale reform (Oakley 1998; O'Connor 2009). Today, more than two thousand sections of the U.S. Code include provisions related to demonstration projects (Kelly and Rocco 2019).

The term *demonstration project* can refer to a heterogeneous mix of governance projects, from pork-barrel spending programs to waivers that allow deviations from statutory law to experimental research projects designed to test the effects of a specific policy intervention (Nathan 2000; Rosenbaum 1992). Despite their diversity, these projects are premised on a common technocratic theory: a program that is carried out in miniature—whose effects are rigorously evaluated—offers policymakers an evidentiary basis for justifying large-scale reform (Cook, Campbell, and Shadish 2002; Nathan 2000). Unlike policy diffusion, in which a policy may spread across jurisdictional boundaries as a result of learning and emulation, demonstration projects are a more formalized experimental process by which policymakers learn about and decide on the initial adoption of a policy within the jurisdictional boundaries of a polity (Karch 2007). Similar to diffusion, however, the “success” of a demonstration does not necessarily lead to learning and program-wide expansion (Gilardi 2010; McCann, Shipan, and Volden 2015; Shipan and Volden 2008). Even when demonstration programs show clear evidence of a policy’s benefits, however, their effects on

politics and policy vary tremendously. On the one hand, some demonstration programs reset the terms of political debate and generate a rapid implementation of new policy instruments (Brodkin and Kaufman 2000; Teles and Prinz 2001). For instance, major changes in Medicare payment—including the development of a prospective payment system—emerged directly from a demonstration undertaken in New Jersey (Cassidy 2008, 9). Within Medicaid, demonstration waivers have also been responsible for enduring programmatic changes, including the introduction of home and community-based services for long-term care, expansions of coverage to new populations, and the emergence of managed care (Thompson 2012, 101–66).

Yet many other demonstrations have failed to alter politics and policy (Cassidy 2008). This is because, even when technical or scientific logics guide the design and execution of projects, they are embedded in a broader institutional and political context (Coyle and Wildavsky 1986). Any policy learning that results from these projects will therefore depend on more than the sum of their treatment effects. One clear example is the Medicare Participating Heart Bypass Demonstration (PHBD). Carried out in the early 1990s, the PHBD illustrated that providing hospitals with a single bundled payment for bypass surgeries led to significant cost savings without sacrificing the quality of care. Despite its statistical success, however, the PHBD faced significant hurdles. First, its expansion depended on congressional approval. This veto point meant that expanding the PHBD required a clear constituency of supporters, but enthusiasm for cost control was limited. Organized medicine remained strongly opposed.¹ Although members of Congress were eager to identify ways of trimming Medicare’s budget, the administrative challenges of implementing the PHBD did not inspire enthusiasm. Further, the delay in releasing the results of the demonstration meant that its supporters missed a critical window of opportunity to expand it during negotiations over the 1997 Balanced Budget Act (Kelly and Rocco 2019).

1. Miriam Laugesen (2016) identifies another case in which organized medicine limited the effectiveness of cost-control initiatives by shaping the construction of the resource-based relative value scale.

As this story reveals, even when demonstration projects produce evidence of statistical success, their expansion into national policies depends on the institutional context in which they are carried out. Institutions may contain veto points that act as a barrier to scaling up even modest reforms. Especially when opposition to a demonstration project is mobilized, interest groups may attempt to leverage the veto points of the congressional or agency-level policy process to prevent the program's expansion into national policy. Institutions can also affect the timing of policy research in consequential ways. As Richard Elmore (1986) notes, policy research often runs on social science time, whereas the time horizon of politics is compressed by budget cycles and elections. Internal rules and procedures that create delay in designing, implementing, and certifying the results of demonstration projects may mean that demonstration research misses critical moments of opportunity for policy change. Finally, institutions can also affect the conditions under which demonstration projects are categorized as a success. Quantitative decision-making tools have the capacity to delimit the scope of participants or choices in debates over public policy (Mennicken and Espeland 2019). Since the late nineteenth century, the development of techniques for measuring actuarial risk, cost-benefit analysis, and the valuation of statistical life have shaped fiscal and regulatory policy in critical ways (Porter 1995; Espeland 1998; Revesz and Livermore 2008; Hood 2017; Elliott 2018; Lakoff and Klinenberg 2010; Power 2004; Muller 2018). These forms of expertise may also constitute an important veto point for the advancement of certain types of policy reforms, whose benefits are not easily calculated or certified using formal methodologies.²

The Patient Protection and Affordable Care Act reshaped the context for the design and expansion of demonstration projects in several fundamental ways. Section 3021 of the ACA expanded and consolidated demonstration authority within a new Center for Medicare and

Medicaid Innovation, earmarking \$10 billion in funding over ten years and \$10 billion in each subsequent decade. CMMI would also be responsible for testing alternative models of value-based payments under Section 3022 of the ACA, which established the new Medicare Shared Savings Program (GAO 2018). Second, the ACA also removed other procedural hurdles to the implementation of demonstration projects. For example, the law prohibited HHS from requiring that demonstration projects be budget neutral, which had impeded the development of effective demonstrations in the past (Kelly and Rocco 2019). It also exempted some elements of the demonstration development process from administrative and judicial review. Most important, the ACA granted the Department of Health and Human Services the authority to expand a Medicare or Medicaid demonstration model on a nationwide basis if the model either reduced spending without reducing the quality of care or improved the quality of care without increasing spending. These effects must be certified by the CMS actuary, and models may not deny or limit coverage or benefit provision. With these new provisions in place, the ACA made it possible for HHS to make significant changes in policy based on the results of demonstration projects *without* seeking additional approval from Congress.

With added discretion to carry out new demonstrations, the Congressional Budget Office projected that CMMI would yield significant savings—reducing federal spending by \$1.3 billion dollars between 2010 and 2019 (CBO 2009). CBO also projected that CMMI demonstrations would contribute to nearly \$5 billion in savings under the Medicare Shared Savings Program. Despite major changes in demonstration authority, CMMI may have had difficulty in accomplishing these objectives for several reasons. First, numerous demonstration models were premised on changes in providers' incentives, organizational culture, and technical infrastructure that might well be difficult to achieve in the span of four or five implementa-

2. The desirability of such analytical veto points may depend in part on how one judges the credibility of the analysis, the commensurability of the policy outcomes in question, or the likelihood that unimpeded programmatic innovation will result in agency loss. Our overriding point, however, is that the presence of these veto points places a brake on rapid programmatic innovation.

tion years (Gore et al. 2020). Second, as CBO has noted, it can be difficult to isolate the effects of certain demonstration projects in a dynamic and complex environment in which multiple institutional changes are occurring simultaneously (Masi and Bradley 2015; Hadley 2016). Third, achieving significant savings depended on the expansion of successful demonstration models nationwide, which required certification by the chief actuary of the Centers for Medicare and Medicaid Services. Yet the actuary was not bound to certify projects for expansion simply because evaluation studies yielded statistically significant findings of cost savings or quality improvement. Rather, the actuarial certification process typically subjected these findings to sensitivity testing and a rigorous review of evidentiary limitations.

In sum, although the ACA may have removed some of the barriers to the expansion of demonstration projects—among them the challenge of convincing Congress to expand successful demonstrations—it may have left other barriers in place and generated a new set of challenges for programmatic innovation. Using demonstration projects to facilitate policy learning requires surmounting not only the challenges of weak institutional capacity and political opposition (Beauregard and Miller 2020), but also technical challenges in commensurating the effectiveness of policy innovation.

DATA AND METHODS

This analysis is based on a qualitative synthesis of data on the implementation of fourteen demonstration models (see table 1). The models are a subset of all demonstration projects completed between the creation of CMMI in 2011 and the end of 2018. Because the outcome of interest is whether a demonstration project is certified by HHS for program-wide adoption, we limit our analysis to demonstration projects that have been completed and evaluated. We therefore exclude the forty-five demonstration projects ongoing at the time of publication. Further, we include only those completed demonstration projects that both began and ended during CMMI's existence. As a result, we ex-

clude demonstrations that began under an earlier authority and institutional context.

In addition, we included only demonstration projects initiated under the authority of ACA Section 3021 or 3022. By considering only those, we are able to exclude demonstration projects directly called for by Congress. For example, we excluded the Community-Based Care Transition Program initiated under Section 3026. By limiting our investigation this way, we include only demonstrations initiated by CMMI. These selection criteria allow us to more fully examine CMMI's autonomous policymaking power. We do, however, exclude two demonstrations that began under Section 3021 authority—the Innovation Advisors Program and the State Innovation Model Pre-Test awards—because neither showed a clear objective to “learn.” The goals of these two programs are, respectively, to create human capital or provide states with financial support to complete a State Health Plan (GAO 2018). Thus neither fits the framework for expansion found in Sections 3021 or 3022. Among the demonstration projects examined here is at least one from each of CMMI's seven categories of innovation models.

To assess the effects of the ACA on the implementation and expansion of demonstration projects, we qualitatively analyzed documents, including published evaluation reports, actuarial certification letters, and program regulations issued pursuant to model findings ($n = 74$ documents, see the online appendix table³). Both authors reviewed all of these documents to examine several features of each demonstration. We first extracted data on the research design, hypotheses, and participants. We next extracted the primary findings contained in each demonstration evaluation report. This included both quantitative and qualitative results indicating whether the demonstration had a statistically significant effect in the hypothesized direction. We collected information on the reported limitations of the evaluation study. Finally, we considered all available evidence on the actuarial certification and expansion of each demonstration, as well as other evidence that demonstration results affected future policy changes within Medicare or Med-

3. Available at <https://www.rsfsjournal.org/content/6/2/67/tab-supplemental>.

Table 1. CMMI Demonstration Models Included in Qualitative Synthesis

Model	Implementa- tion Dates	Participants	Obligations Under 1115A (\$ Millions)	Obligations Under Titles XVIII, XIX (\$ Millions)
Federally qualified health center (FQHC) advanced primary care practice demonstration	2011–2014	434 FQHC sites	64.2	n.a.
Advance payment accountable care organization (ACO) model	2012–2015	35 ACOs	73.8	110.1
Health Care Innovation Awards round 1	2012–2015	108 awardees	967.4	n.a.
Health Care Innovation Awards round 2	2014–2017	39 awardees	497.7	n.a.
Pioneer ACO	2012–2016	32 ACOs	96.9	244.3
Initiative to reduce avoidable hospitalizations among nursing facility residents: phase 1	2012–2016	7 care coordination providers; 143 long-term care facilities	124.7	n.a.
Comprehensive primary care initiative	2012–2016	442 primary care practices	397	0.6
Bundled payments for care improvement (BPCI) model 1	2013–2016	24 hospitals	75.7	n.a.
BPCI model 2	2013–2018	335 hospitals; 204 group practices	see BPCI 1	n.a.
BPCI model 3	2013–2018	758 participants (various facility types)	see BPCI 1	n.a.
BPCI model 4	2013–2018	23 hospitals	see BPCI 1	n.a.
State Innovation Models Initiative, round 1	2013–2016	6 test states, 16 design states	326.7	n.a.
State Innovation Models Initiative, round 2	2014–2017	11 test states, 17 design states	373.7	n.a.
Strong Start for Mothers and Newborns Initiative: enhanced prenatal care models	2013–2018	27 awardees with 200 sites	96.2	n.a.

Source: Authors' analysis based on GAO 2018.

icaid. After reviewing the data from each of the fourteen demonstrations, we each extracted a set of qualitative themes focused on the barriers to, and facilitators of, demonstration implementation, empirical identification of intended results, actuarial certification, and expansion of demonstration models nationwide. We then developed a consensus on themes present across a majority of cases examined.

FINDINGS

As noted, the ACA greatly expands the HHS's authority to expand demonstration projects

without seeking congressional approval and removes several important barriers to project implementation. Nevertheless, several important barriers to expansion remain.

Actuarial Certification and the Challenge of Commensurating Effectiveness

All demonstrations slated for expansion must first be actuarially certified to produce net program savings without reducing quality, or to increase quality without substantially increasing costs. In practice, actuarial certification has been a high bar for demonstration projects to

clear. Indeed, we found that much of the policy learning that CMMI has initiated occurs through means other than the formal expansion of demonstration projects.

To date, only two of CMMI's evaluated models—the Pioneer ACO model and the Medicare Diabetes Prevention Program (MDPP)—have been actuarially certified for expansion. As evidenced by the chief actuary's reports, the standards that guide the certification process are both exacting and somewhat opaque. Even the CBO reports having “limited information” about the standards used to evaluate future demonstrations (Hadley 2016). The MDPP provides a good example here. CMMI funded an initial pilot test of this model through its Health Care Innovation Awards initiative. The model received a positive evaluation, which illustrated that the program resulted in significant improvements in health quality while achieving modest reductions in spending (Hinman et al. 2016). Yet subsequent analysis subjected formal evaluation findings to additional testing, probing the study's limitations with supplemental evidence from clinical trials and the Centers for Disease Control and Prevention's Diabetes Prevention Recognition Program. Finally, the analysis modeled future savings, applying a battery of sensitivity tests (Spitalnic 2016). Although some advocacy organizations later argued that the analysis should have been limited to the formal evaluation report, CMS has been quick to note that the statute makes no such demand (CMS 2016b).

To be sure, evidence of uncertainty did not preclude the model's expansion (Rajkumar 2016). Yet the evaluative standards applied had several important effects on how the MDPP was translated into national policy. Perhaps most important, although the evaluation demonstrated significant reductions in diabetes incidence rates for pre-diabetics without increasing program spending, it was “unclear whether the program would break even over the participants' lifetimes” (Spitalnic 2016, 9). Thus, although the MDPP was initially designed as a lifetime benefit, CMS's expansion ultimately allowed eligible beneficiaries to access MDPP

services only once in their lifetime. Patient advocates pushed back during the rulemaking process, asking CMS to extend the benefit and to allow exemptions from the rule. CMS responded by highlighting that the “once-per-lifetime restriction is necessary in order to generate enough savings to offset the cost of delivering MDPP services” (CMS 2016b, 80470).⁴

The use of break-even analysis in the certification process also exemplifies a broader challenge in scaling up innovations focused on quality improvement. The HHS may expand programs that improve quality, but only if the CMS actuary certifies that expansion “would reduce (or would not result in any increase in) net program spending” (Spitalnic 2015, 1). Actuarial certification thus advantages quality improvements that offset the costs of service delivery. In the MDPP case, the actuary calculated quality improvements into cost savings by modeling the probability that intervention participants would progress from pre-diabetes to diabetes until age eighty-five, multiplying these probabilities by estimated lifetime marginal costs associated with diabetes care, and summing these amounts to produce an expected marginal costs for each intervention starting age (Spitalnic 2016).

Models focused mainly on achieving cost savings without reducing quality, such as the Pioneer ACO, do not typically have to bear this added analytical burden in the certification process (Spitalnic 2015). Instead, they are required to show that quality does not significantly decline in populations attributed to the demonstration. As the MDPP case shows, this constraint on the certification process has not altogether precluded the initiation of demonstration projects focused on quality. Yet it may create incentives for CMMI staff and evaluation researchers to use quality metrics more commensurate with service costs, such as reductions in service use (such as length-of-stay and readmission rates), as opposed to measures that focus primarily on beneficiary mental or physical health status.

Our evidence thus suggests that the process of actuarial certification can preempt the ex-

4. In any case, it is possible that the difficulty of retracting a benefit once it has been made widely available may have informed the agency's decision here (Pierson 1994).

pansion of demonstration projects by making legible statistical uncertainties about their long-term effects on spending or quality. This is obviously true when evaluation study results do not meet the actuary's standards for research design or supplemental evidence is not available to enable modeling or sensitivity testing. Yet even when such evidence is available, as in the case of the MDPP, the actuary's conclusions help determine whether key program attributes—including benefit design and eligibility restrictions—are ultimately expanded.

Policy Learning in the Absence of Actuarial Certification

Although the actuarial certification process posed a barrier to the expansion of demonstration projects into program-wide innovations, it did not preclude CMMI from engaging in policy learning. Our data set includes an average of three complete CMMI-initiated demonstration projects per year between 2014 and 2018. This is in contrast to the pre-ACA period when a growing number of demonstration projects were the result of legislative mandates (U.S. Congress 2009). Because CMMI had been given resources and authority for model development, demonstrations that had no statistical success were not shelved. Instead, they were recalibrated. Even when CMMI models did not become candidates for expansion, they informed policy decisions in more informal ways.

One example of this pattern is the Bundled Payment for Care Improvement Initiative (BPCI) (Lewin Group 2017). Launched in 2013, BPCI aimed to test several models for defining payments and episodes of care in a range of facilities, more than 1,500 in total. In the second iteration of the initiative—model 2—providers were given bonus payments if total Medicare spending for an entire episode of care fell below a target benchmark price set by CMMI. If spending exceeded the target, however, participating facilities had to reimburse Medicare for a share of the spending. In some cases, however, the model evaluation revealed errors in CMMI's target prices. In joint replacement episodes, target prices did not account for spending differences between lower-cost elective surgeries and higher-cost surgeries required after a fracture. As a result of this finding, CMMI de-

veloped a new Comprehensive Care for Joint Replacement (CJR) that deliberately adjusted target prices to account for higher-cost joint replacement surgeries. First-year results from this model indicate that total Medicare spending fell by 3.3 percent more for CJR episodes relative to control-group episodes ($p < .01$) (Lewin Group 2018, 3).

In October 2018, CMMI also announced a new BPCI Advanced model, the goal of which was to expand the bundled-payment concept to a new range of clinical episodes. Under initial BPCI models, CMS reconciled payments with target prices for episodes retrospectively. Although the advanced model is also based on retrospective payment, providers will receive target prices up front “to allow for more effective planning.” Further, CMMI touted new design features intended to elicit physician participation, including streamlined paperwork, risk-adjusted target prices, and “facilitated peer-to-peer learning.” Whether these changes help generate significant savings for Medicare under BPCI Advanced or not, it is clear that CMMI intends them to entice participation. As of April 2019, CMMI expects to enroll 1,299 participants in the model (CMS 2018a, 2018b).

As the BPCI example suggests, CMMI's structure—in a departure from pre-ACA activity—permits the adaptation of existing models in light of ongoing evaluations. Another example of this pattern can be found in the case of the Pioneer ACO. As in other pay-for-performance initiatives, Pioneer featured a two-sided risk model in which ACOs agreed to share savings and losses. Yet Pioneer's model was more aggressive than existing models: savings- and loss-share rates could be as high as 75 percent. Pioneer also gave participants a range of payment agreements with varying degrees of two-sided risk. To incentivize participation in the program, one popular option featured 50 percent one-sided risk in the first implementation year, phasing up to 70 percent two-sided risk in the second year with the option to move to population-based payments in the third (L&M Policy Research 2016b).

The CMS actuary certified the Pioneer model's savings, yet this did not lead to an expansion of the model as such (Spitalnic 2015). It did, however, help generate important changes

in the design of the Medicare Shared Savings Program (MSSP), the keystone of the ACA's effort to shift Medicare payments from "volume to value." In 2016, the Obama administration added a two-sided risk option that included Pioneer's more aggressive 75 percent savings ratio (CMS 2016a). Participation in two-sided risk models within MSSP gradually increased, but 82 percent of MSSP participants still opted for one-sided (CMS 2019). In 2018, the Trump administration further revised the MSSP's rules to engineer ACO participation in two-sided risk models. The new rules reduced the amount of savings available for participation in MSSP's one-sided models by 10 percent and reduced the amount of time ACOs could spend in the one-sided "track" from six years to two years. Moreover, they allowed ACOs in two-sided risk models to offer vouchers or in-kind services to beneficiaries engaging in healthy behavior (CMS 2018b).

Isolating Treatment Effects in Complex Environments

Although the actuarial certification process imposed a new barrier to scaling up demonstration projects, other barriers reflected broader challenges in evaluation research that existed before CMMI, but that were not directly addressed by CMMI's institutional structure and processes. Evaluations of Medicare and Medicaid demonstration projects have long tacitly assumed that the effects of these interventions on spending and care quality can be clearly isolated and identified. Yet in practice these models are implemented in complex environments; multiple policy interventions and market dynamics may complicate the isolation of policy effects. For example, the evaluation of Pioneer ACOs may have understated their cost savings because it leveraged comparisons between patients attributed to the model ACOs and other patients in proximate markets, despite the potential for ACO spillover effects (Spitalnic 2015).

Even though spillover effects in the case of the Pioneer ACO were estimated and adjusted for, other complex dynamics can make it more difficult to evaluate the effects of some CMMI models. One example is the federally qualified health center (FQHC) advanced primary care practice (APCP) demonstration, which was in-

tended to transform FQHCs into patient-centered medical homes (PCMHs). More specifically, the transformation of an FQHC to a Level 3 PCMH required the FQHC achieve specified standards in six categories relating primarily to improving access to and continuity of care, as well as managing and coordinating care (Kahn et al. 2016, 1). The end goal was to transform FQHCs into physician- or nurse-practitioner-directed primary care practices that strive to provide more coordinated, comprehensive, and continuous care to its patients than a standard practice.

To achieve this transformation, CMMI provided participating FQHCs with a care management fee for each eligible Medicare beneficiary served. FQHCs, however, serve a population that goes well beyond Medicare beneficiaries. That Medicare beneficiaries often make up only a minority of an FQHC's patient population created several challenges to the successful completion and evaluation of the demonstration. First, because the care management fees received by FQHCs were determined only by the size of their Medicare population, FQHCs were expected to successfully transform their entire practice with a resource allocation based only on a small portion of their patient population. This payment, which on average amounted to \$6,500 per quarter, was described by many participants as "relatively modest" to achieve the lofty goals of the demonstration (Kahn et al. 2016, xvi). More problematic than a provision of resources that was not commensurate to the design or goals of the demonstration is that nearly all participating FQHCs were receiving additional sources of funding to support the same transformation CMMI was attempting to measure and evaluate. Isolating the effect of the APCP demonstration within the complex policy space occupied by FQHCs was therefore quite challenging for the evaluators. In addition, many of the comparison sites were also receiving technical and financial resources separate from the CMMI demonstration. Researchers' ability to observe significant differences between treatment and control sites was muted by the inability to adequately control the actions of comparison FQHCs given that those sites also sought to transform their practices in the direction of PCMH status.

Infrastructural Gaps and Evidence Translation

Infrastructural gaps constituted barriers to the expansion of several demonstration projects we observed. As Beauregard and Miller (2020) show, gaps in infrastructure or capacity are not unique to CMMI demonstration projects. Yet, in several instances, the implementation of CMMI models was contingent on the adoption of significant changes in data systems, such as those that facilitate patient identification for care management. Yet when participating organizations lacked these systems, generating certifiable evidence of savings within the evaluation period proved difficult.

An illustrative example of this pattern is the Advance Payment (AP) ACO Model, implemented in the spring of 2012. As its title suggests, the AP model aimed to test whether providing an up-front, monthly payment to providers would increase their participation in the Shared Savings Program and whether advance payments would allow ACOs to improve care for beneficiaries, generate Medicare savings more quickly, and increase the amount of Medicare savings. Several months after the model design was finalized, CMS enrolled thirty-six physician-based organizations in the model—roughly 20 percent of all organizations participating in the MSSP. These were, generally speaking, smaller organizations, with the majority having no more than eight thousand beneficiaries per year. CMS furnished participants with an up-front \$250,000 payment, a one-time payment of \$36 per assigned Medicare beneficiary, and monthly payments of \$8 per beneficiary for twenty-four months. Participants spent the majority of these funds on personnel and benefit costs. By contrast, organizations found it difficult to make investments in critical information technology (IT), in part because of higher than expected market prices. The challenge faced by these smaller organizations, particularly in regard to transforming health IT processes and infrastructure, is also visible among the smaller practices that Radhika Gore and colleagues (2020) examine later in this issue. As the results of the AP demonstration would soon show, the challenge of making rapid organizational changes proved to be a significant barrier to accom-

plishing the model's objectives (L&M Policy Research 2016a).

If the goal of the AP model was to demonstrate that up-front payments would quickly generate higher-than-usual levels of savings, the results were decidedly mixed. In 2013, model spending was \$2 per beneficiary per month (PBPM) lower for the AP ACOs than for comparison beneficiaries in the same market not aligned with or assigned to a Pioneer or MSSP ACO. Yet by 2014, AP ACO spending was \$20.80 PBPM higher than expected. Overall, evaluation researchers found that AP ACOs spent "\$70.80 million more in 2014 than would have been spent in the absence of the model" (L&M Policy Research 2016a, ix). Exploratory analysis found that ACOs were more likely to save money if they used claims data or electronic health records to identify patients for care management, or simply had a younger population with fewer chronic conditions.

To be sure, some participants found their experience in the model to be "very powerful," especially when it came to understanding the total cost of care. Yet as the final evaluation report on the model put it, "transforming groups of small, physician-led practices, particularly those with demonstrated need for capital to invest in population management," may simply "take longer than the model period allows" (L&M Policy Research 2016a, 40). Although up-front payments gave organizations the ability to experiment with improving accountability in care, providers likely needed "stronger incentives to reduce overutilization while maintaining quality than they faced under the AP model" (40). Despite the greater bureaucratic discretion and fiscal authority given to CMMI, the interventions it directs often do not provide enough resources or a forceful enough intervention to transform existing models enough to permit scaling.

Another example of this pattern can be found in the case of the Strong Start for Mothers and Newborns Initiative, which aimed to improve maternal and infant health outcomes for women covered by either Medicaid or Children's Health Insurance Program during their pregnancy. To address many perceived weaknesses in prenatal care, the Strong Start Demonstration funded enhanced prenatal services

in three different care models: birth centers, group prenatal care, and maternity care homes. Despite the high social and medical needs of participants, the Strong Start intervention showed significant positive results across nearly every measure of the evaluation. For example, when comparing outcomes among the three models of the demonstration, birth center participants had lower preterm birth rates (4.5 percent) than Group Prenatal Care participants (12 percent) or Maternity Care Home participants (12.9 percent) (Urban Institute 2018, 63). Yet the evaluators were nearly as pessimistic about scalability of the birth center model as they were optimistic about the model's results.

"It is unrealistic," the evaluators concluded, "for Birth Centers to become the dominant maternity care provider under Medicaid or in the U.S. any time soon" (Urban Institute 2018, 144). The larger, system-level barriers to expanding a birth center model include challenges in contracting between birth centers and managed care organizations (MCO), which enrolled 69 percent of Medicaid beneficiaries nationally in 2017, as well as low reimbursement rates paid by MCOs to birth centers (KFF 2017). Medicaid's reimbursements to birth centers outside of MCOs is also identified as a challenge to the expansion of such a model. As a result of payments from Medicaid that are not only too low but also often too slow in arriving, birth centers limit their enrollment of Medicaid-eligible beneficiaries. Beyond the challenges associated with reimbursements and contracting, the scalability of birth center models is constrained by scope-of-practice laws and licensing regulations that limit or completely block midwives from practice. Because such constraints limit the number of birth centers and midwives available to Medicaid beneficiaries, the infrastructure to make scaling such a successful intervention possible is simply not in place. In the absence of broader powers for CMMI to intervene in the complex relationship between payers and providers, CMMI's efforts to implement new and effective approaches to maternal and infant health are highly constrained.

Ostensibly, one near-term effect of some CMMI models and other federal Medicare ini-

tiatives will be to facilitate the development of technical infrastructure where it does not exist. Yet gaps in infrastructure may inhibit participation in these models at the outset. For example, states with lower levels of bureaucratic capacity were significantly less likely to participate in CMMI's State Innovation Models initiative. By contrast, in states with a large number of ongoing demonstration models, the initiative was a way of coordinating and consolidating existing delivery-system reforms (Rocco, Kelly, and Keller 2018). This is to some extent consistent with Beauregard and Miller's (2020) findings on the importance of institutional capacity in states' adoption of 1915(c) waivers within Medicaid.

The Challenge of Limited Time Horizons

A final pattern across our cases was the difficulty of illustrating demonstration model effects in a short period. This was especially true, as noted, when demonstrations require participating entities to first undertake significant efforts to transform and improve physical and technical infrastructure, as well as to alter practice patterns and culture. The Initiative to Reduce Avoidable Hospitalizations for Nursing Facility Residents required participating facilities to dramatically reshape practice culture and processes in the direction of treating more patients in-house rather than transferring those patients to a hospital setting. In other instances, such as the Comprehensive Primary Care demonstration, success hinged on the ability to implement new health IT infrastructures and create competency among staff to use the new capacity. Preparing for a successful implementation, whether through investments in technical infrastructure, adjustments to practice patterns and cultures, or securing buy-in from an array of partners, is time intensive (see Gore et al. 2020). Time, however, is not a luxury that demonstrations often have. Not only do demonstrations operate on tight timelines, often only lasting three or four years, but the necessary preparations and investments must occur while the clock is running. And though CMMI has the ability to iteratively alter demonstration models based on results, the timeline for carrying out these redesigned models remains similarly limited. As a result,

the full power of the intervention is realized for only a small portion of the demonstration period, meaning that any ability to achieve statistically significant, and therefore certifiable, reductions in spending or improved quality is muted.

In the case of the FQHC ACP, the majority of the evaluation period was used to complete the transformation to PCMH status. Achieving level 3 PCMH status was one goal of the demonstration but not the outcome on which the demonstration was ultimately evaluated. The primary outcome of interest, and the criteria on which the demonstration could be certified for program-wide adoption, was whether such a transformation of practice patterns, culture, and technical infrastructure could improve the quality of care and lower the costs of Medicare beneficiaries served by FQHCs. Because the transformation was, itself, such an immense undertaking, most of the FQHCs that succeeded in achieving level 3 status did not do so until the end of the demonstration period. Of the 70 percent of FQHCs that achieved level 3 status, more than half did not achieve this status until the last quarters of the demonstration (Kahn et al. 2016, 21). The final evaluation noted that critical aspects of the intervention, such as the technical assistance given to FQHCs, was not adequately provided until the second year of a three-year demonstration. As a result of such delays, as well as the generally time-intensive process of transformation, little time was available to measure the effect of achieving level 3 PCMH status on the quality and cost of care for Medicare beneficiaries served by FQHCs. In this way, the compressed timelines on which demonstrations are required to operate act as a significant impediment to producing results of the magnitude that could result in certification (Kahn et al. 2016, 249–50).

Perhaps nowhere was time more thematically prominent than in the evaluation of the Initiative to Reduce Avoidable Hospitalizations for Nursing Facility Residents (the initiative). The initiative, which was a partnership between Enhanced Care and Coordination Provider (ECCP) organizations and long-stay nursing facilities, was aimed at reducing the costs and improving the health and health care of nurs-

ing facilities' residents by reducing the number of unnecessary hospitalizations. Each demonstration site was partnered with an ECCP organization whose nurses provided educational and direct clinical interventions of varying intensity (RTI 2017, 12–15). Although achieving health information technology (HIT) implementation and competency on a compressed timeline was challenging for the initiative much as it was for the CPC, the most important and time-consuming transformation was not infrastructural, but rather cultural and procedural. Indeed, as dictated by the design of the demonstration, in which the nursing facilities were partnered with ECCP nurses, the primary intervention shared across all demonstration sites was the transfer of knowledge from the ECCP nurses to nursing facility staff and providers. This knowledge transfer was largely responsible for the cultural and procedural shifts that were required for success. "Implementing an initiative of such scope requiring a shift in facility culture and adjustments to care processes," the evaluators concluded, "take a significant amount of time and cannot be achieved quickly" (RTI 2017, ES-18).

Participants in this demonstration were quick to acknowledge that the likelihood of returning positive effects within the designated time frame was slim. Interviews with ECCP leadership noted that the four-year demonstration window was inadequate to show observable effects on costs and quality (RTI 2017, 238). In the state-by-state evaluations, the slow, gradual process of transformation was directly noted in multiple lessons learned sections. It was not surprising, then, that the results produced in 2014 were considerably weaker than those produced in 2015 (80). Interviewees, however, did express confidence that additional time would likely produce measurable, positive findings in regard to reducing avoidable hospitalizations and lowering costs. Indeed, many participants noted their enthusiasm for continuing the initiative's efforts as participants in the subsequent Payment Reform Initiative. That many of the initiative's participants would continue in a related demonstration shows both the promise of the demonstration to improve quality and reduce costs as well as the challenges of achiev-

ing such results in a single demonstration period. In this way, a demonstration that is unable to return certifiable results over the course of its four- or five-year window can act as a runway that launches a subsequent demonstration. Under Section 3021(c), HHS may expand the duration and scope of demonstration projects, which can facilitate this type of iterative learning process.

The CPC demonstration is another illustrative example of the difficulty of returning certifiable results within a demonstration's short operational window. This demonstration also shows how one demonstration can jumpstart and improve the likelihood of a subsequent and related demonstration returning more positive and measurable outcomes. In the case of the CPC, a demonstration that evaluators described as a "bold undertaking," 78 percent of payers that remained in the demonstration until its conclusion later joined the subsequent CPC+ demonstration. In addition, 98 percent of practices participating in CPC later joined CPC+ (Mathematica Policy Research 2018, xlv). The CPC's final evaluation directly noted how the results and lessons learned from the CPC informed the development of the CPC+ model. Such a pattern, in which the majority of a demonstration's participants are selected for and enroll in a subsequent and related demonstration is itself a recognition that the operational resources and capacity developed over one demonstration window is not often sufficient for producing certifiable results (10).

DISCUSSION

In this article, we analyze the ACA's effects on the federal government's ability to diffuse innovations in Medicare and Medicaid payment and delivery policies. In the past, expanding successful demonstration projects required congressional approval and faced numerous procedural barriers, such as severe budget-neutrality requirements (Kelly and Rocco 2019). The ACA eliminated many of these roadblocks, consolidated demonstration authority, and gave HHS the ability to expand successful demonstration projects, subject to actuarial certification. Evidence from early demonstration projects suggests that these statutory changes have facilitated iterative learning. When mod-

els fail to produce statistically significant savings or quality improvements, CMMI's expanded authority and capacity enables policymakers to learn from these findings, redesigning payment parameters, incentives, and policy infrastructure as necessary in future models. The availability of a larger suite of ongoing models has also enabled policymakers to make important adjustment to larger value-based purchasing initiatives within the Medicare program.

Yet though the ACA provided federal officials extensive authority to expand successful demonstration models without seeking congressional approval, other important barriers to expansion remain. Essentially, demonstration projects continue to face the challenges associated with commensurating and making legible programmatic success. On the one hand, several of the identified challenges are common to demonstration research writ large. Even prior to CMMI, limited time horizons, the difficulty of isolating treatment effects, and inadequate capacity have made it difficult to scale up successful demonstration projects (Kelly and Rocco 2019). Yet CMMI's approach to implementing demonstrations did not fundamentally surmount those challenges. Moreover, even though the ACA removed the barrier of congressional approval, its actuarial certification requirements created a barrier of their own. Indeed, even when demonstration projects yield strong evidence of statistically significant savings, the CMS Office the Actuary subjects these findings to sensitivity tests and modeling exercises that make legible statistical uncertainties contained in traditional evaluation studies. This is fully in keeping with its role as a guardian of program integrity. Nevertheless, it has narrowed the types of demonstration projects that can be considered for formal expansion, potentially limiting CMMI's capacity to generate the program savings initially projected by CBO.

Of course, it may also be that CBO's projections were simply overly optimistic given the scale of the challenges faced with the transition from volume to value. The inability to generate savings at the CBO projected level is not an instance of a policy "designed to fail" or the product of Republican sabotage (Levy, Ying, and

Bagley 2020), but instead a result of a policy design that leaves certain procedural hurdles to innovation in place while erecting new ones. As our qualitative synthesis reveals, the effects of many CMMI demonstration models would take longer to observe than the short evaluation windows typically afford. Although models often aim to affect spending through narrow changes in payment parameters, they often demand broader cultural and technological changes in health-care facilities that cannot be accomplished quickly. In some cases, providers simply lack the technical infrastructure that would truly facilitate these changes. Even when this is not the case, the complexity of the implementation environment can make it difficult to recover evidence of models' effectiveness through standard techniques of evaluation research. In other words, although innovation—however one defines it—might be happening in CMMI models, difference-in-differences models are not necessarily the ideal instruments for registering it.

It seems clear, then, that the institutional apparatus for implementing demonstration projects affects not only whether successful interventions are expanded into national policy, but the kinds of interventions that are deemed successful (Espeland 1998; Hood 2017; Porter 1995; Revesz and Livermore 2008; Elliott 2018; Lakoff and Klinenberg 2010; Power 2004; Muller 2018). How one judges the ACA's effects on payment and delivery-system reform may depend on her orientation toward interbranch politics and the specific set of policy ideas advantaged and disadvantaged by this process. On the one hand, congressional skepticism toward the executive branch and a demand for program integrity have made actuarial certification a desirable procedural safeguard, its effects on policy innovation notwithstanding. By contrast, policymakers in Westminster-style systems tend to craft vague statutes and judges tend to show more deference to executive interpretation of these statutes (Moe and Caldwell 1994; Kelemen 2009). In such a system, it is conceivable that organizations like CMMI would experience fewer procedural constraints.

On the other hand, as the ACA's authors clearly knew, the goals of legislative control and program integrity are to some extent in conflict

with the idea of executive-led programmatic innovation. Actuarial certification has no doubt made it more difficult for HHS to expand potentially cost-saving or quality-improving reforms nationwide. Further, the need to undertake significant transformations on a compressed timeline has sidelined innovations that require long-term technical or cultural changes to accomplish. Along these lines, CMMI has often favored applicants for demonstration models with existing technical abilities or those that have already begun the desired transformation. If the barriers to actuarial certification and program-wide adoption remain prohibitively steep, and CMMI's transformative powers remain largely operative at the micro-level, such selection processes may further exacerbate inequalities within the health-care delivery system. Both the micro-level transformations and the concerns regarding health-care inequalities are evident in the study of IMPACT by Gore and colleagues (2020). Although the authors highlight beneficial organizational transformations facilitated by IMPACT, they also raise concerns about the challenges faced by small, immigrant-serving practices when participating in demonstration projects. At present, however, the ACA's structure trades off the rapid diffusion of innovations in favor of cautious actuarial judgment. Whether this choice will ultimately serve the goal of moving from "volume to value" remains to be seen.

None of this is to suggest that CMMI has not helped stimulate innovation and policy learning. Indeed, the qualitative sections of many demonstration evaluation reports noted substantial, positive transformations in practice patterns and culture, as well as improved infrastructure and care delivery. The FQHC demonstration, for example, despite not producing actuarially certifiable results, succeeded in facilitating the transition of 70 percent of participating FQHCs to level 3 PCMH status. Other demonstrations, such as the CPC, provided not only important lessons on model design to subsequent demonstrations, but also initial infusions of resources and capacity on which those subsequent demonstrations would build.

A lingering question, then, is how innovative we should expect CMMI to be given the hand it has been dealt. Indeed, it might be that,

through carrying out a broad range of interventions, CMMI is slowly generating the conditions for a transformed health system in a way that is difficult to track through formal approaches to evaluation research and short-term observational windows. Nevertheless, our results suggest that policymakers should exercise caution both in designing programs intended to generate innovation and setting expectations regarding when, where, and how their effects will be most readily observed.

REFERENCES

- Appleby, Julie. 2016. "Obamacare's Test Kitchen for Payment Experiments Faces an Uncertain Future." NPR, November 30. Accessed June 30, 2019. <https://www.npr.org/sections/health-shots/2016/11/30/503863311/health-law-s-test-kitchen-for-payment-changes-could-offer-tool-for-gop-ideas>.
- Beauregard, Lisa Kalimon, and Edward Alan Miller. 2020. "Why Do States Pursue Medicaid Home Care Opportunities? Explaining State Adoption of the Patient Protection and Affordable Care Act's Home and Community-Based Services Initiatives." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 6(2): 154–78. DOI: 10.7758/RSF.2020.6.2.07.
- Brodkin, Evelyn Z., and Alexander Kaufman. 2000. "Policy Experiments and Poverty Politics." *Social Service Review* 74(4): 507–32.
- Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862–1928*. Princeton, N.J.: Princeton University Press.
- Cassidy, Amanda. 2008. *The Fundamentals of Medicare Demonstrations*. Background Paper no. 63. Washington, D.C.: The George Washington University. Accessed December 26, 2019. http://www.nhpf.org/library/background-papers/BP63_MedicareDemos_07-22-08.pdf.
- Centers for Medicare and Medicaid Services (CMS). 2016a. "Medicare Program; Medicare Shared Savings Program; Accountable Care Organizations—Revised Benchmark Rebasing Methodology, Facilitating Transition to Performance-Based Risk, and Administrative Finality of Financial Calculations." *Federal Register* 81, no. 112 (June 10): 37950–8017.
- . 2016b. "Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule and Other Revisions to Part B for CY 2017; Medicare Advantage Bid Pricing Data Release; Medicare Advantage and Part D Medical Loss Ratio Data Release; Medicare Advantage Provider Network Requirements; Expansion of Medicare Diabetes Prevention Program Model; Medicare Shared Savings Program Requirements." *Federal Register* 81, no. 220 (November 15): 80170–562.
- . 2018a. "CMS Announces Participants in New Bundled Payment Model." October 9. Accessed March 31, 2019. <https://www.cms.gov/newsroom/press-releases/cms-announces-participants-new-value-based-bundled-payment-model>.
- . 2018b. "Medicare Program; Medicare Shared Savings Program; Accountable Care Organizations—Pathways to Success and Extreme and Uncontrollable Circumstances Policies for Performance Year 2017." *Federal Register* 83, no. 249 (December 31): 67816–8082.
- . 2019. "Fast Facts Archives." Accessed March 31, 2019. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/shared-savingsprogram/Downloads/fast-facts-archives.zip>.
- Congressional Budget Office (CBO). 2009. "Cost Estimate for the Amendment in the Nature of a Substitute to H.R. 3590, as Proposed in the Senate on November 18, 2009." Washington: Government Printing Office. Accessed March 31, 2019. <https://www.cbo.gov/publication/41423>.
- Cook, Thomas D., Donald Thomas Campbell, and William R. Shadish. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, Mass.: Houghton Mifflin.
- Coyle, Dennis J., and Aaron Wildavsky. 1986. "Social Experimentation in the Face of Formidable Fables." In *Lessons from Income Maintenance Experiments*, edited by Alicia Munnell. Boston, Mass.: Federal Reserve Bank of Boston.
- Elliott, Rebecca. 2018. "'Scarier Than Another Storm': Values at Risk in the Mapping and Insuring of US Floodplains." *British Journal of Sociology* 70(3): 1067–90. DOI: 10.1111/1468-4446.12381.
- Elmore, Richard F. 1986. "A Political Scientist's View of the Income Maintenance Experiments." In *Lessons from Income Maintenance Experiments*, edited by Alicia Munnell. Boston, Mass.: Federal Reserve Bank of Boston.
- Espeland, Wendy Nelson. 1998. *The Struggle for Wa-*

- ter: *Politics, Rationality, and Identity in the American Southwest*. Chicago: University of Chicago Press.
- Gilardi, Fabrizio. 2010. "Who Learns from What in Policy Diffusion Processes?" *American Journal of Political Science* 54(3): 650–66.
- Gore, Radhika, Ritu Dhar, Sadia Mohaimin, Priscilla M. Lopez, Anna A. Divney, Jennifer M. Zanowiak, Lorna E. Thorpe, and Nadia Islam. 2020. "Changing Clinic-Community Social Ties in Immigrant-Serving Primary Care Practices in New York City: Social and Organizational Implications of the Affordable Care Act's Population-Health-Related Provisions." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 6(2): 264–88. DOI: 10.7758/RSF.2020.6.2.12.
- Government Accountability Office (GAO). 2018. "CMS Innovation Center: Model Implementation and Center Performance." GAO-18-302. Washington: Government Printing Office. Accessed March 31, 2019. <https://www.gao.gov/assets/700/690875.pdf>.
- Guterman, Stuart, Karen Davis, Kristof Stremikis, and Heather Drake. 2010. "Innovation in Medicare and Medicaid Will Be Central to Health Reform's Success." *Health Affairs* 29(6): 1188–93.
- Hadley, Mark. 2016. "CBO's Estimates of the Budgetary Effects of the Center for Medicare & Medicaid Innovation." Testimony before the Committee on the Budget, U.S. House of Representatives, September 7. Washington: Congressional Budget Office. Accessed December 26, 2019. <https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/reports/51921-cmmitestimony.pdf>.
- Hinnant, Laurie, Sima Razi, Rebecca Lewis, Amy Sun, Martia Alva, Tom Hoerger, Sara Jacobs, and Michael Halpern. 2016. "Evaluation of the Health Care Innovation Awards: Community Resource Planning, Prevention, and Monitoring, Annual Report 2015: YMCA of the USA." Research Triangle Park, N.C.: RTI International. Accessed December 26, 2019. <https://innovation.cms.gov/Files/reports/hcia-ymcadpp-evalrpt.pdf>.
- Hood, Katherine. 2017. "The Science of Value: Economic Expertise and the Valuation of Human Life in US Federal Regulatory Agencies." *Social Studies of Science* 47(4): 441–65.
- Kahn, Katherine L., Justin W. Timbie, Mark W. Friedberg, Peter Mendel, Liisa Hiatt, Emily K. Chen, Amii M. Kress, Christine Buttorff, Tara A. Lavelle, Beverly A. Weidmer, Harold D. Green, Mallika Kommareddi, Rosalie Malsberger, Aaron Kofner, Afshin Rastegar, and Claude Messan Setodji. 2016. *Evaluation of CMS's Federally Qualified Health Center (FQHC) Advanced Primary Care Practice (APCP) Demonstration: Final Report*. RR-886/2-CMS. Santa Monica, Calif.: RAND Corporation. Accessed December 26, 2019. <https://downloads.cms.gov/files/cmmti/fqhc-finalevalrpt.pdf>.
- Kaiser Family Foundation (KFF). 2017. "Total Medicaid Managed Care Enrollment." Accessed December 26, 2019. <https://www.kff.org/medicaid/state-indicator/total-medicare-mc-enrollment>.
- Karch, Andrew. 2007. *Democratic Laboratories: Policy Diffusion Among the American States*. Ann Arbor: University of Michigan Press.
- Kelemen, R. Daniel. 2009. *The Rules of Federalism: Institutions and Regulatory Politics in the EU and Beyond*. New York: Oxford University Press.
- Kelly, Andrew S., and Philip Rocco. 2019. "From 'Trial and Error' to Major Reform: The Politics of Medicare Demonstration Projects." *Public Administration* 97(3): 621–38. DOI: 10.1111/padm.12585.
- Lakoff, Andrew, and Eric Klinenberg. 2010. "Of Risk and Pork: Urban Security and the Politics of Objectivity." *Theory and Society* 39(5): 503–25.
- Laugesen, Miriam J. 2016. *Fixing Medical Prices: How Physicians Are Paid*. Cambridge, Mass.: Harvard University Press.
- Levy, Helen, Andrew Ying, and Nicholas Bagley. 2020. "What's Left of the Affordable Care Act? A Progress Report." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 6(2): 42–66. DOI: 10.7758/RSF.2020.6.2.02.
- Lewin Group. 2017. *CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 3 Evaluation and Annual Monitoring Report*. Falls Church, Va.: The Lewin Group. Accessed December 26, 2019. <https://downloads.cms.gov/files/cmmti/bpci-models2-4yr3evalrpt.pdf>.
- . 2018. *CMS Comprehensive Care for Joint Replacement Model: Performance Year 1 Evaluation Report*. Falls Church, Va.: The Lewin Group. Accessed December 26, 2019. <https://innovation.cms.gov/Files/reports/cjr-firstannrpt.pdf>.
- L&M Policy Research. 2016a. *Evaluation of CMMI Accountable Care Organization Initiatives: Advance Payment ACO Final Report*. November 25. Washington, D.C.: L&M. Accessed December 26,

2019. <https://innovation.cms.gov/Files/reports/advpayaco-fnevalrpt.pdf>.
- . 2016b. *Evaluation of CMMI Accountable Care Organization Initiatives: Pioneer ACO Final Report*. December 2. Washington, D.C.: L&M. Accessed December 26, 2019. <https://innovation.cms.gov/Files/reports/pioneeraco-finalevalrpt.pdf>.
- Masi, Paul, and Tom Bradley. 2015. "Estimating the Budgetary Effects of Legislation Involving the Center for Medicare & Medicaid Innovation." *CBO blog*, July 30. Accessed December 26, 2019. <https://www.cbo.gov/publication/50692>.
- Mathematica Policy Research. 2018. *Evaluation of the Comprehensive Primary Care Initiatives: Fourth Annual Report*. Baltimore, Md.: Mathematica Policy Research. Accessed December 26, 2019. <https://downloads.cms.gov/files/cmimi/CPC-initiative-fourth-annual-report.pdf>.
- McCann, Pamela J. Clouser, Charles R. Shipan, and Craig Volden. 2015. "Top-Down Federalism: State Policy Responses to National Government Discussions." *Publius: The Journal of Federalism* 45(4): 495–525.
- Mennicken, Andrea, and Wendy Nelson Espeland. 2019. "What's New with Numbers? Sociological Approaches to the Study of Quantification." *Annual Review of Sociology* 45(1): 223–45. DOI: 10.1146/annurev-soc-073117-041343.
- Moe, Terry M., and Michael Caldwell. 1994. "The Institutional Foundations of Democratic Government: A Comparison of Presidential and Parliamentary Systems." *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft* 150(1): 171–95.
- Muller, Jerry Z. 2018. *The Tyranny of Metrics*. Princeton, N.J.: Princeton University Press.
- Nathan, Richard. 2000. *Social Science in Government: The Role of Policy Researchers*. Albany, N.Y.: Rockefeller Institute Press.
- Oakley, Ann. 1998. "Public Policy Experimentation: Lessons from America." *Policy Studies* 19(2): 93–114.
- O'Connor, Alice. 2009. *Poverty Knowledge: Social Science, Social Policy, and the Poor in Twentieth-Century US History*. Princeton, N.J.: Princeton University Press.
- Orren, Karen, and Stephen Skowronek. 2017. *The Policy State: An American Predicament*. Cambridge, Mass.: Harvard University Press.
- Pierson, Paul. 1994. *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment*. New York: Cambridge University Press.
- Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J.: Princeton University Press.
- Power, Michael. 2004. *The Risk Management of Everything: Rethinking the Politics of Uncertainty*. London: Demos.
- Rajkumar, Rahul. 2016. "Request for Determination of Eligibility for Expansion of Diabetes Prevention Program." March 14. Baltimore, Md.: Centers for Medicare & Medicaid Services. Accessed December 26, 2019. <https://innovation.cms.gov/Files/x/mdpp-certification.pdf>.
- Revesz, Richard L., and Michael A. Livermore. 2008. *Retaking Rationality: How Cost-Benefit Analysis Can Better Protect Our Environment and Our Health*. New York: Oxford University Press.
- Rocco, Philip, Andrew S. Kelly, and Ann C. Keller. 2018. "Politics at the Cutting Edge: Intergovernmental Policy Innovation in the Affordable Care Act." *Publius: The Journal of Federalism* 48(3): 425–53.
- Rosenbaum, Sara. 1992. "Mothers and Children Last: the Oregon Medicaid Experiment." *American Journal of Law and Medicine* 18(1): 97–126.
- RTI. 2017. *Evaluation of the Initiatives to Reduce Avoidable Hospitalizations Among Nursing Facility Residents: Final Report*. Walham, Mass. RTI International. Accessed December 26, 2019. <https://downloads.cms.gov/files/cmimi/irahnr-finalevalrpt.pdf>.
- Shipan, Charles R., and Craig Volden. 2008. "The Mechanisms of Policy Diffusion." *American Journal of Political Science* 52(4): 840–57.
- Shrank, William. 2013. "The Center for Medicare and Medicaid Innovation's Blueprint for Rapid-Cycle Evaluation of New Care and Payment Models." *Health Affairs* 32(4): 807–12.
- Spitalnic, Paul. 2015. "Certification of Pioneer Model Savings." April 10. Baltimore, Md.: CMS Office of the Actuary. Accessed December 26, 2019. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/ActuarialStudies/Downloads/Pioneer-Certification-2015-04-10.pdf>.
- . 2016. "Certification of Medicare Diabetes Prevention Program." March 14. Baltimore, Md.: CMS Office of the Actuary. Accessed December 26, 2019. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research>

- /ActuarialStudies/Downloads/Diabetes-Prevention-Certification-2016-03-14.pdf.
- Teles, Stephen M., and Timothy S. Prinz. 2001. "The Politics of Rights Retraction: Welfare Reform from Entitlement to Block Grant." In *Seeking the Center: Politics and Policymaking at the New Century*, edited by Martin Levin, Marc Landy, and Martin Shapiro. Washington, D.C.: Georgetown University Press.
- Thompson, Frank J. 2012. *Medicaid Politics: Federalism, Policy Durability, and Health Reform*. Washington, D.C.: Georgetown University Press.
- Urban Institute. 2018. *Strong Start for Mothers and Newborns Evaluation: Year 5 Project Synthesis*, vol. 1: *Cross-Cutting Findings*. Washington, D.C.: Urban Institute. Accessed December 26, 2019. <https://downloads.cms.gov/files/cmmi/strongstart-prenatal-finalevalrpt-v1.pdf>.
- U.S. Congress. House of Representatives. 2009. "America's Affordable Health Choices Act. Report of the House Energy and Commerce Committee on H.R. 3200." 111th Cong. 1st sess. H Rept. 111-299, pt. I.