

Marquette University

**e-Publications@Marquette**

---

Political Science Faculty Research and  
Publications

Political Science, Department of

---

9-2019

## **From 'Trial and Error' to Major Reform: The Politics of Medicare Demonstration Projects**

Andrew S. Kelly

Philip B. Rocco

Follow this and additional works at: [https://epublications.marquette.edu/polisci\\_fac](https://epublications.marquette.edu/polisci_fac)



Part of the [Political Science Commons](#)

---

Marquette University

**e-Publications@Marquette**

***Political Science Faculty Research and Publications/College of Arts and Sciences***

***This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript.*** The published version may be accessed by following the link in the citation below.

*Public Administration*, Vol. 97, No. 3 (September 2019): 621-638. [DOI](#). This article is © Wiley and permission has been granted for this version to appear in [e-Publications@Marquette](#). Wiley does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Wiley.

# From 'Trial and Error' to Major Reform: The Politics of Medicare Demonstration Projects

Andrew S. Kelly

Department of Health Sciences, California State University, California

Philip Rocco

Department of Political Science, Marquette University, Milwaukee, Wisconsin

## Abstract

Facing fragmented institutions and partisan polarization, officials in the United States often attempt to engineer policy change without assembling new legislative majorities. To this end, they have increasingly employed demonstration projects, policy innovations undertaken by administrative agencies designed to test alternative approaches to implementation or service delivery on a limited segment of the target population and for a limited period of time. Despite the increasing importance of demonstration projects, they are an undertheorized source of policy change. In this article, we conceptualize demonstration projects as part of a class of experimental institutions that, while incremental in scope, have the potential to 'scale up' into more substantial reforms. Data from three

Medicare demonstrations suggest that policy change is more likely when programmes generate strong support constituencies; minimize administrative and infrastructural costs; are undertaken in contexts with few veto points; and align with the time horizons of elected officials.

## INTRODUCTION

In government systems with large numbers of veto points, major policy reform is a rare event (Immergut [42]; Steinmo and Watts [81]). Even when significant changes are possible, they are imperiled by forces of ideological polarization, inadequate implementation resources and weak institutional design (Patashnik and Zelizer [68]). In the face of these constraints, officials in the United States rely on institutional structures that permit gradual policy change without the need to assemble legislative majorities (Hacker [38]; Mahoney and Thelen [53]; Orren and Skowronek [67]). One such structure is the demonstration project—a policy innovation undertaken by executive-branch agencies designed to test alternative approaches to implementation or service delivery on a limited segment of the target population and for a limited period of time. In recent years, Congress has given executive agencies (and sometimes state governments) limited authority to waive existing programmatic rules in order to test drive new policy ideas addressing problems as diverse as highway construction, poverty, and healthcare (Oakley [64]; Thompson [87]). For example, three-quarters of states now run their Medicaid programmes as demonstration projects, which account for more than a third of total annual Medicaid expenditure (Government Accountability Office [33], p. 1).

Despite their increasing importance, demonstration projects are an undertheorized tool for authorizing gradual institutional change in the United States (Mahoney and Thelen [53]). On the one hand, demonstration projects resemble a piecemeal and partial form of authority. While they allow entrepreneurial bureaucrats or presidential administrations to experiment with ideas that would not be legislatively viable, they are restricted in geographical, technical, or fiscal scope, and may be easily eliminated by future administrations (Orren and Skowronek [67], p. 98). Public officials often note significant political challenges in translating successful demonstration projects into broad-based reforms (Wilensky [97]). At the same time, empirical studies have illustrated how demonstration projects ignite long-lasting changes in governing authority (Teles and Prinz [86]; Cassidy [16]). One reason for this confusion is that policy literature often treats demonstration projects as technocratic sites of 'experimentation' (Shadish et al. [79]). Yet these projects often have explicitly political origins and effects (Coyle and Wildavsky [18]; Brodtkin and Kaufman [10]; Nathan [63]).

In this article, we conceptualize demonstration projects as part of a class of experimental institutions that, while incremental in scope, have the potential to 'scale up' into more substantial policy changes. To better understand demonstration projects as sites of gradual institutional change, we draw on insights from the literature on policy learning and experimentation. We highlight four features of demonstration projects that we expect to be associated with more significant policy change or 'scaling up'. To probe the plausibility of this theory, we examine three demonstration projects undertaken within the Medicare programme, two of which 'scaled up' into national reform and one of which did not. This article is, therefore, an effort at both theory building and theory testing, goals which are particularly well served by a case-based methodological approach. We leverage both within- and cross-case analyses using primary and secondary source material, as well as key informant interviews (George and Bennett [31]; Mahoney and Thelen [54]). In recognizing the strengths and limitations of

case-based analyses, we seek only to understand the general conditions under which a demonstration project 'scales up' and to generate insights that can be tested across different policy areas and different experimental institutions. To this end, we conclude with a discussion of the results and limitations of our study, as well as the implications of our findings for future research.

## DEMONSTRATION PROJECTS, EXPERIMENTAL INSTITUTIONS, AND THE POLITICS OF SCALING UP

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 [67]). Beginning in the late nineteenth century, executive-branch agencies developed deliberately 'experimental' programmes that tested new approaches to public service delivery (Carpenter [13]). By the 1930s, Congress had begun to use the term 'demonstration project' to refer to intergovernmental grants designed to stimulate the creation of state-level public resources (Wallis [95]). The expansion of government during the 'long Great Society' dramatically expanded the use of demonstration projects in order to test controversial ideas in social policy (O'Connor [66]). The 1980s saw a significant increase in the number of congressional bills, both introduced and enacted, containing demonstration provisions (see Figure). This trend has ebbed in recent years, yet over 2,000 sections of the US Code currently contain provisions relating to demonstration projects.

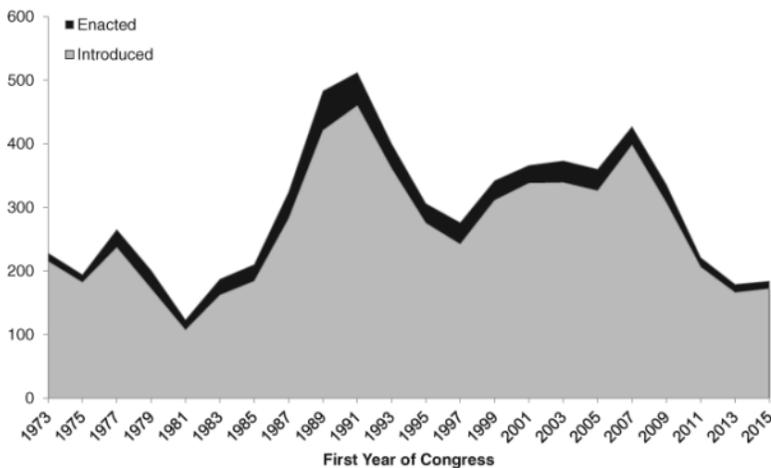


Figure 1. Congressional legislation containing demonstration provisions, 1973–2016 Source: Authors' analysis of Congressional Record

While officials' reliance on demonstration projects has arguably increased, they remain undertheorized as a source of gradual policy change (Hacker [38]; Mahoney and Thelen [53]). One reason for this is that the term 'demonstration project' refers to a heterogeneous mix of governance projects. This includes distributional (i.e., 'pork barrel') spending programmes (Nathan [63], p. 37), waiver programmes that allow federal or state agencies to deviate from existing law (e.g., demonstration projects under Section 1115), as well as a range of research projects—governmental and non-governmental alike—that systematically evaluate the effects of a specific policy intervention through a formalized research process, such as randomized controlled trials (Rosenbaum [75]; Marino et al. [55]). It should also be noted that the published literature on demonstration projects tends to focus on

strategies for technical and methodological refinements rather than the politics of major reform and experimental institutions (Nathan [63]; Shadish et al. [79]; Granger and Maynard [34]; Peck [71]).

Unsurprisingly, then, existing research draws a variety of inconsistent conclusions about the effects of demonstration projects on major policy reform. On the one hand, demonstration projects are easily characterized as weak sources of significant policy change. In many cases, even when demonstration projects illustrate that new policies are technically feasible and consistent with the values of governing officials, they may fail to instigate significant changes in public policy (Coyle and Wildavsky [18]; Nathan [63]). By contrast, Brodtkin and Kaufman ([10]) argue that demonstration projects might serve as 'shadow institutions' that re-set the terms of policy debate. During the 1980s, 'welfare to work' demonstrations in the states changed the political dynamic of welfare reform, 'setting in place a new political and policy path' that undermined political support for traditional welfare benefits (Teles and Prinz [86], p. 224). Major changes in Medicare payment—including the development of a prospective payment system—emerged directly from a demonstration undertaken in New Jersey (Cassidy [16], p. 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 [87], pp. 101–66).

Another reason for this confusion might be that US scholars have overlooked the growing European research literature on the politics of policy learning (see Dunlop and Radaelli [25] for a review). This literature has revealed a class of experimental institutions, similar to demonstration projects, in which policy-makers test the effects of alternative policy designs. These institutions comprise a range of initiatives—from NGO-led pilots of conditional cash transfer programmes (Peck and Theodore [70]) to government-led randomized controlled trials of interventions aimed at controlling bovine tuberculosis (Dunlop [23], [24]).

Generally speaking, experimental institutions provide a potential arena for translating small-scale execution of policy into large-scale policy change. Yet under what conditions does this translation occur? Building on the observation that policy experiments rarely, if ever, follow a simple problem-solving process, some scholars identify the accessibility of experimental evidence to policy-makers as a critical factor in understanding how and when experimental institutions influence policy outcomes (Jowell [45]; Stoker and John [83]; Stoker [82]). The accessibility of experimental evidence is determined by the form in which the evidence is presented, its relevance and, most importantly, its timeliness. As Stoker ([82], p. 53) rightly points out, there is often a disjuncture that exists between the linear experimental process and the decidedly non-linear political process. Experimenters, it is argued, must exploit 'windows of opportunity' if they are to influence policy-making decisions.

In contrast to this focus on the experimental 'toolkit', Dunlop ([23]) raises the possibility that because some research methods (e.g., randomized controlled trials) are likely to help policy-makers 'clinch' their argument, they become uniquely susceptible to contestation and deconstruction at the hands of interest groups, policy experts, and elected officials. Yet, despite the importance of political interests, ideas, and institutions that shape the outcomes of experimental institutions, these variables are too easily lumped into broad concepts such as 'power or politics' or used as an error term to explain away the 'chaos' or messiness of the learning or experimental process (Stoker [82], p. 55).

As Dunlop's work suggests, there is a need for greater attention to how political forces that surround experimental institutions shape their capacity to scale into national policy. Scaling up policy experiments often involves a fundamentally political battle over who has the authority to speak (Blyth [ 9]; Dunlop and Radaelli [26]). Hence, some recent work has highlighted how institutions affect the power of technocratic actors to shape the range of available policy solutions when policy anomalies or crises emerge (Matthijs and Blyth [57]). As Campbell and Pedersen ([11], pp. 74–77) show, when political institutions fragment authority for knowledge production, as in the US case, there are limited opportunities for consensus formation on the definition of policy problems and solutions. Second, past policy choices can affect the learning process by influencing policy-makers' access to and perception of information about the effects of specific reforms (Béland [ 7]). Finally, institutions can also help to socially construct how policy-makers view 'failures' and 'successes' (Blyth [ 9]; Schmidt [78]). Policy-makers' tacit acceptance of neoliberal assumptions about the appropriate role of the state and the market may cause them to ignore or discount evidence about policy effectiveness that challenges these background ideas (Carstensen and Matthijs [15]).

Drawing on this literature, we argue that four factors affect the likelihood that demonstration projects will scale up into broad-based reforms. The first two factors concern how demonstration projects are designed. First, demonstration projects may be more likely to 'scale up' when they mobilize broader coalitions of support. As Dennis J. Coyle and Aaron Wildavsky ([18], p. 182) note: 'experimentation may point the way toward specific policy solutions once there is sufficient consensus to make broad support possible, but research cannot replace the dialogue among supporters of different ways of life'. Perhaps unsurprisingly, then, in the face of strong interest-group opposition, Congress has failed to expand numerous Medicare demonstration projects despite strong statistical evidence of effectiveness (Cassidy [16]). In contrast, demonstration projects that receive support from cross-cutting constituencies may be more likely to scale up. Such constituencies might attach themselves to demonstration projects because of their interest in promoting specific policy instruments or pursuing policy goals that the project aims to secure (Béland and Howlett [ 8]). Hence it may be the case that demonstration projects that pursue a variety of goals or employ a variety of instruments will generate cross-cutting constituencies. Diversified goals not only increase the likelihood of finding valuable interest-group support; they also limit the risk that a single null or adverse finding will damage the reputation of the project.

Second, demonstration projects may be more likely to scale up when their expansion imposes relatively low administrative costs and requires little investment in implementation infrastructure. This may be the case when demonstrations do not unnecessarily conflict with pre-existing policy legacies or bureaucratic priorities. Alignment with existing policy legacies may help to assure risk-averse policy-makers that the transition costs associated with expanding the demonstration will be minimal. For instance, bureaucrats may structure demonstration projects to rely on widely available and cheap technologies or strong existing implementation networks.

We also argue that there are two contextual factors that affect the likelihood that demonstration projects will 'scale up'. First, we argue that demonstration projects are more likely to scale into national reforms in institutional contexts with few veto points. Especially when mobilized opposition to a demonstration project exists, interest groups may attempt to leverage the veto points of the

congressional or agency-level policy process to prevent the programme's expansion into national policy. Thus when congressional approval is required in addition to agency approval, we predict that the expansion of demonstration projects will be more easily stalled by concentrated interests (Baumgartner et al. [6]). Interest groups certainly do possess leverage over policy decisions made at the agency level (Yackee and Yackee [99]; McKay and Yackee [58]); however, when agencies have the authority to carry out expansions of successful demonstration projects in the absence of congressional approval, interest-group opposition may be less relevant because in such instances, interest groups have fewer potential veto points to activate.

Finally, we argue that timing matters. American politics operates on a compressed time horizon of annual budget cycles, biennial congressional elections, and quadrennial presidential elections. By contrast, policy research often runs on 'social science' time (Elmore [27]). Significant policy experiments may take a decade or more to unfold from design to analysis and critical appraisal. Hence, it may not be possible for data generation, collection, and analysis to be sped up or slowed down to provide policy ideas or solutions in response to a sudden, pressing problem, a change in administration, or a shifting national mood that opens what Kingdon ([50]) describes as a 'window of opportunity'. The following sections explore each of these propositions in the context of Medicare demonstration projects.

## DEMONSTRATION PROJECTS AND THE POLITICS OF MEDICARE TRANSFORMATION

We examine the implementation of demonstration projects undertaken within the Medicare programme. Demonstration projects in Medicare have their origins in the fiscal crisis the programme experienced during its first decade of operation (Ball [5]). This crisis led to the creation of general demonstration authority for cost control under the 1972 Social Security Amendments (Zarabozo [100]). Since the 1980s, the Center for Medicare and Medicaid Services (CMS), formerly the Health Care Financing Administration (HCFA), has conducted numerous cost-saving demonstration projects of its own, as well as those mandated by statute (Cassidy [16]). Even so, it is rare for Congress to give HCFA or CMS the authority to expand a demonstration based on evaluation findings. As a high-level Medicare official explained to us in an interview, even if the evaluation findings were purely positive, the agency's Office of Legislation would have to develop a proposed bill, go to Congress, and generate political energy (Interview, 14 December 2017). In 2010, the Patient Protection and Affordable Care Act (ACA) significantly increased CMS's authority to expand Medicare demonstrations into national policy. Under the changes brought about by the ACA, CMS has the authority to expand a demonstration into a programme-wide policy without the approval of Congress when a demonstration yields evidence of actuarially certified decreases in cost without a reduction in quality or improvements in quality without increases in cost (Guterman et al. [37]).

Medicare demonstrations provide a fertile ground for building theories about the processes and politics of change within experimental institutions (Greenwald [36]). Despite abiding congressional interest in expanding demonstrated payment reforms, provider opposition to many of these proposed changes often remains strong. Indeed, as one Medicare official told us, the process of demonstration can occasionally give opponents of the proposed reform time to mobilize, especially when programme

evaluations unfold over a long period of time (Interview, 15 January 2018). Thus the problem is not simply the asynchronicity of political and social science time, but the political opportunity structure created by long-term research projects. Efforts to scale demonstrations must also clear executive-level hurdles, chief among them being the Office of Management and Budget (OMB). Stringent fiscal criteria imposed by the OMB have led to the termination of demonstrations that promised to reduce overall government spending, but could not promise reductions in each programme involved. Agency administrators can push back on these criteria to protect high-priority programmes, but the OMB's criteria have nevertheless prevented demonstration projects from being developed in the first place. Unsurprisingly, a key internal criterion of success for officials at CMS is whether demonstration projects can 'get off the ground' at all (Interview, 15 December 2017). In short, the design of demonstration projects involves a delicate balance between concerns about fidelity to the principles of evaluation research and sensitivity to fiscal and political criteria that must be met in order for the project to 'scale up'.

Our analysis relies on three case studies of Medicare demonstration projects, one that was ultimately not expanded into national policy (the Participating Heart Bypass Demonstration, or PHBD) and two that were (the Physician Group Practice Demonstration, or PGPD, and the Medicare Diabetes Prevention Program, or DPP) (see summary in Table). For each case, we collected all relevant evaluation studies (see list in online appendix). Using ProQuest, we reviewed all relevant hearings, legislative action, and press coverage related to each demonstration, starting one year prior to its initiation and ending one year after its final evaluation report was published. Finally, between 2017 and 2018, we conducted three semi-structured interviews with federal officials and non-government actors involved in the execution of Medicare demonstration projects. All interviewees were recruited by email and all were former high-level federal officials from the Centers for Medicare and Medicaid Services (CMS), as well as with CMS's non-government partners, including officials that served during both Democratic and Republican administrations and across the entire time period covered by this study. Key informant interviews of this nature provide a source of causal process observations that are uniquely able to elucidate the decision-making process at the highest levels of policy-making. We are, for example, able to see how intergovernmental relationships constrain the consideration of policy options, how interest group activities are understood and perceived by government officials, and how seemingly technocratic actors participate in a highly political environment.

Table 1. Summary of cases

	<b>Participating Heart Bypass Demonstration</b>	<b>Physician Group Practice Demonstration</b>	<b>Diabetes Prevention Program</b>
Mobilized opponents	Yes	Yes	No
Mobilized supporters	No	Yes	Yes
Administrative / infrastructure costs required for scaling up	High	Minimal	Minimal
Congressional approval required for scaling up?	Yes	Yes	No

Time between beginning of formal demonstration project and first results	10 years (1988–98)	9 years (2000–09)	3 years (2013–15)
Outcome	Expansion proposal died in Committee	Expanded version included in ACA	Expanded via HHS decision

## The political failure of the Participating Heart Bypass Demonstration

The Tax Equity and Fiscal Responsibility Act (TEFRA) of 1982 marked a significant shift in the politics of Medicare: providers now faced government-imposed limits on reimbursement. The 1980s brought a wave of similar changes, in the form of prospective payments for acute-care hospitals through Diagnostic Related Groups (DRGs) and physician pay freezes (Oberlander [65]). Yet while these changes helped to control hospital payments under Medicare Part A, physician services paid under Medicare Part B continued to grow (Davis and Burner [20]). This was in part because providers shifted services that were previously provided under Part A to Part B. Moreover, all hospital inputs (even long ICU stays) were available 'free' to physicians, who bore none of the financial risk when expensive procedures required additional Part A spending. At the same time, technological changes and advertising campaigns created demand for expensive, and often unnecessary, new procedures.

Heart bypasses were a hallmark of this trend. By the late 1980s, bypass surgeries had come under public scrutiny as an expensive and overused procedure; one prominent study demonstrated that nearly 30 per cent of all heart bypasses, angioplasties, and other treatments were unnecessary (Flanigan [30]). Even so, between 1985 and 1988 Medicare payments for bypass surgery grew by 12 to 14 per cent. HCFA officials now faced both congressional demand for cost control and cross-pressure from physician groups like the American Thoracic Society to reimburse what they believed were valuable, life-saving procedures (Rich [74]; Millenson [61]).

To address this problem, HCFA officials designed the Participating Heart Bypass Demonstration (PHBD) to incentivize hospitals to engage in cost-control measures. The mechanism of choice was the so-called 'bundled' (or global) payments for heart bypass surgeries, in which providers are paid one lump-sum for the episode of care, rather than a payment for each individual service and provider involved. The discounted payment—negotiated with participating hospitals and physicians—was meant to give hospitals and physicians incentives to reduce the costs of care while maintaining high standards. As HCFA administrator Gail Wilensky framed it for reporters, 'Physicians and hospitals are pushed to work together. They have a single pot of money and it's got to cover everything' (Stout [84]). Hospitals would then offset discounts with improved efficiency (lower overhead) and increases in the volume of procedures. HCFA officials claimed they could advance this goal by allowing hospitals to drive volume by marketing themselves directly to consumers as federal 'centres of excellence' for heart bypass services (Rosenblatt [76]).

If the PHBD was to be declared a success, HCFA officials first had to show that payment bundling saved money overall. Measured by this standard, they succeeded. From the start of the programme to its completion in 1996, Medicare saved \$42.3 million on bypass patients treated in demonstration hospitals. This was a 10 per cent decrease in expected spending on bypass patients; 86 per cent of the savings came from HCFA-negotiated discounts on inpatient payments. In addition, beneficiaries and

insurers saved \$8.9 million in Part B coinsurance payments, which HCFA rolled into a total estimated \$50.3 million savings in five years (Cromwell et al. [19], pp. 1–20).

Yet despite the PHBD's statistical success, several aspects of the programme threatened its political viability. First, the costs of implementing and expanding the demonstration were high. No pre-existing infrastructure was in place to support payment bundling, and the experience of both hospitals and HCFA officials with negotiating bundling agreements was limited (Cromwell et al. [19]). Instead, HCFA officials essentially had to tailor the procedure to participating hospitals as the demonstration evolved. This highlighted a broader problem with expanding payment bundling systemwide: bundles were especially unresponsive to rapid advances in technology that altered the price of services (Herdman et al. [41]). Further, HCFA was in a weakened bargaining position with providers. In addition to a congressional prohibition on exclusive contracting with hospitals, the agency lacked the technical capacity to quickly verify the discount rates offered by participating hospitals. This involved a cumbersome and error-laden process of manually linking Part B physician bills with Part A hospital bills. Missing data tortured cost analysts. For some claims, no Part B bills could be found that corresponded to Part A bills. HCFA also puzzled over how to calculate non-included physicians. These contingencies required additional face-to-face negotiations before rates could be agreed upon (Cromwell et al. [19]).

These technical challenges greatly delayed the roll-out of the demonstration, which meant that results were not available for congressional review during an important window of opportunity. By the time that Jerry Cromwell's team at Health Economics Research, Inc. put the finishing touches to the demonstration's final evaluation report in 1998, Congress had passed the Balanced Budget Act of 1997 (BBA), which aimed to save nearly \$400 billion over 10 years through cuts to fee-for-service payments and premium increases, rather than payment reforms (Kahn and Kuttner [46]). With the passage of the BBA, the agenda shifted and decision-makers were no longer casting about for the solution that Jerry Cromwell's team now had in hand. The window of opportunity had closed.

Second, the PHBD made enemies within organized medicine. In 1993, an AMA spokesperson rejected the idea entirely, claiming that it would create 'incentives to underserve patients', blur the distinction between physician and hospital services, and allow hospitals to exert 'undesirable pressures on the provision of physician services to patients' (Marwick [56]). Initially, hospitals and some physician groups, including the American Thoracic Society, appeared to be key support constituencies (Interview, 15 January 2018). HCFA's pitch to participating hospitals was that they might increase their market share by implementing a federal 'Centre of Excellence' label in their advertising. Yet improvements in market share for participating hospitals failed to materialize. Among the four original participants, only two significantly increased their market share. For all hospitals in subsequent cohorts, market share actually declined (Cromwell et al. [19], pp. 1–19).

In addition, HCFA promised hospitals that the implementation of bundled payments would help to achieve significant overhead reductions. On the surface, the demonstration achieved this goal, yet because PHBD imposed new financial risks on their physicians, participating hospitals were forced to expend additional resources on achieving provider buy-in. At some hospitals, surgeons leveraged their bargaining power to refuse to accept payment rollbacks. In other cases, hospitals pre-empted controversy by providing surgeons with bonuses and other in-kind payments such as more operating-

room time and payment for physicians' assistants. Such moves kept physicians 'on board' with the demonstration, but resulted in fiscal and administrative challenges for hospitals (Cromwell et al. [19]).

Organized groups mobilized to maintain the status quo on Medicare payment policy (Baumgartner et al. [6]). At a 1997 hearing of the House Ways and Means Committee, an AMA representative testified before the Senate Finance Committee that physicians were concerned that the PHBD had not demonstrated adequate cost savings and quality improvements, and that 'Centers of Excellence' would ultimately become 'Centers of Cost Cutting' (US Congress, Ways and Means [91], p. 78). The PHBD received no support from hospitals or other significant stakeholder groups. And because of the delay in publishing the results, HCFA was also left without a clear defence. Proposals for a 'centre of excellence' approach mobilized additional opposition among conservative think tanks. Stuart M. Butler, then a vice-president of the conservative Heritage Foundation, rebuked the idea for appearing to be anti-competitive. As Butler put it, allowing the federal government to determine centres of excellence would be like 'an umpire in the World Series working for one of the teams' (Pear [69]).

Thus by the time that Senator Bob Graham (D-FL) introduced a proposal to expand the 'centres of excellence' model in the Senate Finance Committee, it was quickly shot down 18 to 2, with Republicans leading the charge against it (Pear [69]). Indeed, while the PHBD yielded promising results, congressional approval was ultimately required to expand the programme into national policy. Yet the initial costs of roll-out, misalignment between the programme's time horizons and those of elected officials, and persistent opposition from core interest groups made it unlikely that Congress would adapt Medicare policy based on the programme's results.

#### P4P politics and the Physician Group Practice Demonstration

The experience of managed care in the 1990s made an indelible mark on providers and consumers, leaving both groups sceptical of care models they believed would limit access to medically necessary treatments (Mechanic [59]). Despite some favourable evidence from the private sector, Medicare HMOs also showed little ability to effectively control costs. The backlash to managed care did not ultimately change the cost-control paradigm that had defined Medicare debates throughout the 1990s (Jacobson and Jones [44]). By 2000, cost-control advocates had strong incentives to learn about alternative policy instruments that could both address their fiscal concerns about fee-for-service and minimize the backlash that had met managed care (Gold [32]). The solution came in the form of demonstration projects that tested several 'pay for performance' (P4P) schemes that had been developed by business coalitions and philanthropic foundations in the late 1990s. P4P programmes were diverse but relied on a basic logic: 'in order to improve quality (and hopefully lower costs), provider behavior must be measured against performance indicators and compensated accordingly' (Tanenbaum [85], p. 720).

P4P thus reflected a growing attention to quality (Institute of Medicine [43]). Unlike the Medicare demonstrations of the 1990s, the goal of P4P projects was not cost control alone, but improvements in pre-specified quality metrics. This shift had an important effect on the politics of payment reform. Unlike cost control, which tended to unify provider organizations, the issue of quality acted as a wedge between them. Whereas the AMA initially opposed performance-based approaches, other groups—including specialist organizations like the American College of Physicians—ultimately agreed that P4P arrangements were consistent with their principles. In 2005, after an internal debate within the AMA's

House of Delegates, the organization ultimately agreed to develop 140 measures in 34 clinical areas. CMS would use these measures in P4P demonstration projects (Tanenbaum [85]).

Arguably the most influential P4P design was the 'shared savings' model, under which providers were able to keep the savings they generated for Medicare in exchange for implementing new quality measures and engaging in performance evaluation activities. While former HHS Secretary Louis Sullivan had helped to promote such models in the early 1990s, the AMA strongly opposed legislative proposals to develop them into full-scale demonstration projects (Levy and Borowitz [52]). Subsequent studies of interdisciplinary care by researchers at Brandeis University led to the inclusion of one such project, the Physician Group Practice Demonstration (PGPD), in the Benefits Improvement and Protection Act of 2000 (Tompkins et al. [88]; Wallack and Tompkins [94]). Under the programme, participating physician groups were eligible to keep a portion of the savings they generated for Medicare relative to a pre-set spending target, in exchange for assessing their performance on 32 quality measures. If they improved performance on measures, their share of savings increased (Kautter et al. [48]).

The results of the PGPD were promising. Over the course of the first three years, six of the 10 groups shared in approximately \$46 million in savings. By beating their expenditure targets, five PGPs generated Medicare savings of \$38.7 million, earning performance payments of \$31.7 million. The programme's evaluation report attributed these successes to strong organizational integration, physician leadership and expertise with non-Medicare payer initiatives, as well as the adoption of robust health IT infrastructures (Kautter et al. [47]). Thus even before the final results were published, the Medicare Payment Advisory Commission (MedPAC), the independent agency advising Congress on issues of access, quality, and cost within the Medicare programme, had begun to use the PGPD as a key model for the concept of Accountable Care Organizations (ACOs) (Miller [62], p. 12).

In 2009, congressional deliberations over expanding the PGPD into the Medicare Shared Savings Program (MSSP) also highlighted two elements of the demonstration that enabled swift expansion. First, when compared with bundled payments, the design of the PGPD integrated more seamlessly with existing policy infrastructure. In testimony before the House Ways and Means Committee in 2009, Dartmouth's Elliott Fisher, who had pioneered the ACO concept, praised the demonstration. In his words, expanding the PGPD into a nationwide reform would 'require little disruption of current practice and referral patterns within most markets within the United States, meaning that almost all physicians and hospitals could participate in such networks' and would not require 'any change to current fee-for-service reimbursement' (Fisher [29], pp. 3–4). This was in part because the demonstration built on 'natural referral networks' that provided a 'large proportion of the care to their Medicare beneficiaries'. As a result, it could be expanded without requiring patients to be 'locked into specific providers' (Fisher [29], p. 8). Testimony by the Urban Institute's Robert Berenson also highlighted how the PGPD allowed professionals in small practices to collaborate with physicians as part of 'virtual' teams, with 'chronic care support activities residing in the community'. This approach, he said, would be 'more practical for those parts of the country which are unlikely to form multispecialty or large single-specialty groups' (US Congress, Ways and Means 2009, p. 52).

Second, the PGPD's focus on quality changed the coalitional dynamics surrounding the demonstration. Whereas the provider community was largely opposed to government cost-control measures as such, issues of value and quality tended to fracture support between the 'old guard' within the AMA and

state medical societies, who remained sceptical, and specialist groups who more readily accepted the need for reforms to ensure better performance on quality and safety metrics (Tanenbaum [85]). Representing the old guard at congressional hearings were physicians like Todd Williamson, the President of the Medical Association of Georgia. As Williamson put it to the members of a House subcommittee, 'physicians have always embraced the value of working as a team when providing care', yet his society strongly disagreed that 'basing payment on participation in a "team" created by the government is the appropriate model for payment and delivery systems' (Williamson [98], p. 7). By contrast, the American College of Surgeons did not directly oppose the expansion of the PGPD, but instead urged that 'when measuring the quality and cost of care' in ACOs, 'it will always be critical to risk-adjust to account for possible complications and outcomes' (US Congress, Ways and Means 2009, p. 199).

The PGPD also garnered support from organizations representing consumers. Whereas cost-control demonstrations generated fears of coverage denials, consumers embraced the demonstration's focus on quality metrics. The president of the AARP (formerly the American Association of Retired Persons) Jennie Chin Hansen, praised the Senate Finance Committee's decision to expand PGPD as one that would 'reward high quality care rather than how much care is provided' (Hansen [40], p. 7). A group representing non-retired consumers, Health Care for America Now, also argued that the committee should embrace the PGPD's approach to quality measurement in order to enhance patient safety (Kirsch [51], p. 14).

While the PGPD took nearly a decade to roll out, the release of its results coincided with the prioritization of health reform as a policy issue by congressional Democrats, who maintained unified control of government during the 111th Congress. By the time congressional deliberations on the PGPD began, MedPAC reports already included the PGPD as a key example of how to design a voluntary ACO, in part owing to its promising findings on quality improvement on quality measures in the four key areas identified by the demonstration: diabetes, congestive heart failure, coronary artery disease, and preventive care (Medicare Payment Advisory Commission (MedPAC) [60], p. 49). To be sure, MedPAC acknowledged that the PGPD still had significant issues with cost-savings. Yet as its report argued, these problems were artifacts of the demonstration's focus on quality. The most plausible explanation for the findings on cost was that the PGPD's quality metrics required better detection and coding of patient illnesses than occurred at comparison sites (MedPAC 2009, p. 49). Increases in routine screenings led to better quality scores, but they also increased the likelihood that patients would receive higher risk scores, leading to increases in cost growth. Thus when members of a health subcommittee of the House Energy and Commerce Committee highlighted the PGPD's effects on cost growth, MedPAC chairman Glenn Hackbarth emphasized that such errors were to be expected given the demonstration's greater attention to quality. Indeed, the MedPAC chairman suggested that such results only highlighted the need for expanding the demonstration. As Hackbarth put it, 'On an idea like ACOs, we are unlikely to get it exactly right the first time, so there needs to be ongoing cycles of refinement and improvement. That requires discretion and resources' (US Congress, House Committee on Energy and Commerce 2009, p. 644).

With MedPAC's imprimatur, a troika of House committees soon finalized language expanding the PGPD into an ACO pilot programme 'to allow physicians in small- and mid-sized practices to form an ACO

without disrupting care for their patients' (US Congress, House Budget Committee [89], p. 417). In justifying the expansion, the report on the House Reconciliation Bill acknowledged witness testimony showing that the PGPD would allow Medicare to move from 'simply paying physicians for the volume of care they provide toward paying for the value of care delivered' (US Congress House Budget Committee [89], p. 417). And while the Committee acknowledged concerns raised about the possibility that providers would underprovide needed care in attempting to meet their savings targets, 'no evidence of this behavior has been observed under the PGP demonstration' (US Congress House Budget Committee [89], p. 418). The demonstration is now widely acknowledged to have set the course for ACOs under the Patient Protection and Affordable Care Act (Evans [28]).

### Lay organizations and the expansion of the Diabetes Prevention Program pilot

In 2016, direct medical spending related to diabetes approached \$176 billion annually, prompting the Secretary of Health and Human Services, Sylvia Burwell, to describe the disease as a financial burden on both families' and the nation's finances (Carey [12]). For Medicare, in particular, diabetes presents a dire set of challenges. There are nearly 11 million seniors, or roughly 20 per cent of the Medicare population, who suffer from diabetes (American Medical Association [3]). The average cost of treating a Medicare beneficiary with diabetes is \$15,700 annually (Ratner et al. [73]). Medicare, as indicated by the \$15,700 annual price tag, has historically financed diabetes treatment, but has not included benefits aimed at diabetes prevention.

Historically, the addition of new Medicare benefits has been a steep political and policy challenge. Described as a 'negative consensus', Medicare's stability has applied almost equally to its liberal principles and general funding levels as to its benefit structure (Oberlander [65], p. 139). Medicare's sustainability does not mean that the programme has been frozen, but the enactment of new benefits has only occurred on rare occasions, with coverage for End-Stage Renal Disease (1972) and prescription drugs (2003) being the primary examples. Failed legislative attempts to add diabetes prevention to Medicare have dotted the programme's recent policy history. Given the difficult path for legislatively driven benefit expansions, the certification of the Diabetes Prevention Program (DPP) for nation-wide expansion in 2016 was greeted with considerable enthusiasm by prevention advocates and Medicare supporters.

The Medicare DPP pilot has its roots in the 1990s. Between 1996 and 1999, the Diabetes Prevention Program Research Group conducted a randomized clinical trial that demonstrated that an intensive lifestyle intervention was more effective at lowering the incidence of diabetes than the use of a pharmaceutical, metformin (Diabetes Prevention Program Research Group [22]). At the end of the trial, the lifestyle intervention, which included 16 one-hour face-to-face meetings with individual participants, lowered the incidence of diabetes by 58 per cent, compared to 31 per cent with the use of metformin (Carroll [14]). In what would eventually make the trial of particular interest to Medicare, the incidence of diabetes among persons age 60 or older decreased by 71 per cent with the lifestyle intervention (Carroll [14]).

Despite the strong positive results, questions remained as to the costs and feasibility of deploying such an intervention on a large scale. The existing infrastructure to operate such an intensive intervention was quite limited, while the costs associated with an intervention that relied on one-on-one sessions were prohibitively high. In 2003, to address such challenges, the intervention was redesigned and

deployed through the YMCA (now known as the Y) (Carroll [14]). The 16 interventions were redeveloped to function in a group setting and to be run by employees of the Y. The intervention, however, was still not widely adopted in the fight against diabetes. In 2009, a partnership was formed between the Y, UnitedHealth Group (the nation's largest private insurer), and the Centers for Disease Control that expanded the Y-based diabetes intervention to 43 states. It was this iteration that provided the research and rationale for the Diabetes Prevention Act of 2009. Finally passed as part of the much larger Affordable Care Act, the Diabetes Prevention Act appropriated money to the CDC for the National Diabetes Prevention Program (NDPP).

In 2012, the Center for Medicare and Medicaid Innovation (CMMI) granted the Y \$11.8 million to operate a diabetes prevention pilot for Medicare beneficiaries (Carey [12]). The DPP pilot tested the feasibility and effect of expanding the DPP to the Medicare population as a whole. Enrollees attended weekly meetings that instructed them in dietary change, increased physical activity, and other behavioural changes aimed at reducing the risk of developing type 2 diabetes. The results showed strong health gains and significant cost reductions. The Office of the Actuary at CMS determined that Medicare could save \$2,650 for each enrollee in the Diabetes Prevention Program over 15 months (CMS Office of the Actuary [17]).

Opposition to Medicare reform runs highest when proposed changes raise the spectre of imposing losses on beneficiaries (Peterson [72], p. 154). The addition of a new benefit may not initially seem to fit alongside controversial proposals to increase the eligibility age or increase cost-sharing, but long-sought reforms have previously come under fire when they have created a real or perceived threat of cost increases or a reduction of other benefits (Saldin [77]). Like the P4P reforms, there was an incentive to learn how to expand benefits in a manner that did not increase costs or cause a reduction in the quality or extent of coverage in other areas. This is a difficult policy needle to thread, but demonstrating such an ability, even on a small scale, would go a long way in mobilizing supporters and demobilizing opposition. The DPP pilot showed that the programme-wide adoption of the DPP pilot would establish a new benefit, reduce spending, improve quality, and would do so without limiting coverage or benefits in other areas. As a result, proponents of the DPP were able to minimize the real and perceived fear that the creation of a new benefit could equate to a net loss for beneficiaries.

Like the P4P demonstration, the DPP pilot was part of the larger movement to reward value over volume in Medicare. Importantly, the DPP was not focused solely or even primarily on cost control, but on improving quality through an expansion of benefits. Still, the DPP was projected to save Medicare \$1.3 billion over ten years. Despite the overall savings, the DPP is estimated to produce \$7.7 billion in new Medicare spending in the form of reimbursements for preventative services that were not previously covered. Included among the eligible providers are laypersons who are certified as prevention practitioners. A design feature of this kind had the potential to become a source of conflict between traditional providers and the prevention practitioners, mobilizing traditional providers and interest groups like the AMA against the DPP. Under the pilot's structure, however, primary care providers are still required to perform all initial screenings and identify potential qualifying beneficiaries. Under the design's structure, primary care providers are also required to perform follow-up services to determine the effectiveness of the lifestyle interventions (US Congress, Finance Committee [93]). These services, even in the absence of performing the intervention itself, represent a

new source of revenue for traditional Medicare providers. 'There's money in DPP and prevention', one person involved with the DPP stated, 'so people are diving in' (Interview, 12 January 2018).

Because prevention-focused programmes commonly struggle to create a support constituency (Khatana et al. [49]), it was particularly important for the DPP's viability to mobilize support, or at least demobilize opposition. The focus on quality improvements and the establishment of a new set of reimbursable preventive services helped mobilize advocacy organizations like the American Diabetes Association as well as interest groups like the AMA. In addition, affected interest groups like the AMA were involved by the demonstration's designers and implementers during the implementation process (Interview, 12 January 2018). Describing the DPP pilot as a 'groundbreaking effort' the AMA urged CMS to include diabetes prevention in Part B coverage (American Medical Association [4]). Similarly, the American Diabetes Association applauded what it described as the 'landmark' and 'historic' announcement that the DPP would be covered under Medicare (American Diabetes Association [2]).

The DPP's incorporation of laypersons and community-based organizations into its delivery infrastructure proved influential for two reasons. First, the incorporation of a broader universe of reimbursable providers helped mobilize a more diverse support coalition. In addition to the existing universe of Medicare providers, the coalition came to include community-based organizations like the Y, as well as for-profit corporations like Omada Health and Weight Watchers, all of whom would be certified as Medicare-reimbursable providers. The congressional lobbying efforts of corporate actors like Omada Health and Weight Watchers included direct calls for Congress to expand Medicare coverage to include the DPP and for CDC-certified laypersons to be eligible for Medicare reimbursements (Weight Watchers [96]). The Diabetes Advocacy Alliance, which counts among its members both Omada Health and Weight Watchers, strongly urged the Senate Finance Committee's Chronic Conditions Working Group to allow entities that are not currently providers under Medicare to provide the DPP benefit (Diabetes Advocacy Alliance [21]). In a crowded policy environment, a policy's prospects for being placed on the agenda and advanced toward enactment can receive a boost from prominent and loud cheerleaders. The DPP's design ensured that its cheerleaders would include traditional Medicare providers and advocacy organizations, as well as non-profit community organizations and for-profit corporations.

The use of laypersons in the provision of the lifestyle intervention was also critical to lowering the price of the intervention. In its earliest iterations, the high cost of the intervention made scaling the programme prohibitively expensive. When the intervention was provided through 16 one-on-one sessions, the cost of the intervention was \$1,476 per person (Carroll [14]). In the redesigned intervention that was operated in a group setting by the Y, the cost fell to \$205 per person. The lower cost of administering the intervention was among the factors that allowed proponents to tout figures like the \$1.3 billion saved over ten years and the \$2,650 reduction in costs for Medicare beneficiaries. Senator Chuck Grassley (R-IA), a key Republican member of the bipartisan group supporting diabetes prevention, cited cost savings side-by-side with quality improvements in his statement supporting CMS's proposed expansion: 'Preventing [diabetes] is important to a person's health and quality of life. It's also important to controlling Medicare costs' (Grassley [35]). Grassley's policy preferences for both Medicare cost control and benefit expansion were satisfied by the goals and focus of the DPP.

The reliance on laypersons also followed past policy decisions and utilized an existing infrastructure. In authorizing a National Diabetes Prevention Program, the ACA jump-started the construction of an even wider infrastructure for the delivery of the intervention. In 2017, for example, there were more than 1,300 organizations registered to deliver NDPP-like interventions across all 50 states (Ackermann [1]). This growth, according to one of the intervention's original architects, 'underscores the incredible growth in our nation's capacity to provide DPP-like interventions' (Ackermann [1]). The infrastructure that was constructed by this ACA-authorized effort was based on DPP standards and the CDC criteria for certifying prevention providers. In building upon an existing infrastructure, the DPP pilot reduced both the operational and political costs of the demonstration and its subsequent expansion. Just as employing laypersons in providing the intervention helped lower costs, the use of an existing national infrastructure of community-based organizations and virtual counsellors reduced expenditures by precluding the time-consuming and costly task of building new operational and infrastructural capacity. The Y's infrastructure and nationwide footprint made it an ideal partner for scaling the DPP. With 10,000 programme sites across the United States, the Y was well positioned to not only identify solutions that could improve health, but was also uniquely capable of scaling such solutions with an existing infrastructure. The Y, as one person involved in the demonstration described it, is a 'scaling engine' (Interview, 12 January 2018).

When HHS announced the decision to expand the DPP pilot programme-wide, there were already close to 800 CDC-certified prevention programmes in existence (Khatana et al. [49]). Relying on this existing infrastructure of CDC-certified programmes reduced the need to invest in additional infrastructure and further made scaling more likely by reducing the administrative costs by relying, at least in part, on lower cost laypersons. Furthermore, the use of an existing physical and human infrastructure of providers created a broad coalition of support among the numerous actors already occupying the diabetes prevention policy space. If, for example, the designers of the DPP pilot embraced the view that the intervention should be provided under the direction of a physician rather than by certified laypersons, Omada Health, the largest CDC-recognized provider of the DPP, would have likely become an opponent rather than a leading private sector proponent. The design of the DPP pilot, and the decision to build upon past policies, therefore, created friends rather than enemies among the existing stakeholders and reduced the costs and time required to expand the programme system-wide. Finally, because the decision to scale the DPP did not require congressional approval, there were fewer potential veto points for any opponents to activate, making the decision to scale more likely.

## CONCLUSION

Demonstration projects are an increasingly prevalent form of policy-making in the United States. Yet they remain undertheorized as a source of gradual policy change, portrayed contrastingly as weak instruments of incrementalism and powerful 'shadow institutions' that re-set the terms of political debate. As our results suggest, neither of these images captures the process by which demonstration projects lead to significant changes in public policy. Rather, demonstration projects are part of a class of experimental institutions in which policy-makers test the effects of alternative policies. Their capacity to generate policy change is thus conditional on a set of political and institutional factors which our case studies illuminate.

These case studies illustrate that not all statistically successful demonstration projects are created equal. By most accounts the PHBD was successful in reducing unnecessary medical procedures, improving quality, and lowering Medicare costs. Yet it faced a number of institutional and political barriers to expansion. Not only did it fail to mobilize potential supporters, it ignited a blaze of criticism from well-resourced opponents. The administrative costs and infrastructure required to scale the programme into national policy did not win it many adherents in Congress. Contextual factors accentuated these problems: in the pre-ACA period, expanding the PHBD would have required congressional approval. Delays in the roll-out of the demonstration itself, however, meant that it missed a crucial window of political opportunity in which the ideas or solutions sought by decision-makers matched those provided by the PHBD. The cases of the PGPD and the DPP further emphasize that demonstrations with stronger support coalitions and low 'scaling' costs are more likely to scale into national policy, especially if the institutional context contains few veto points (e.g., congressional approval) and project roll-out is synchronous with the time horizons of elected officials.

While these political factors play a critical role in determining whether policy-makers will scale demonstration projects, such factors are not typically foremost in the minds of the individuals who design and implement them. Our interviews revealed that programme-level administrators viewed the technical success of demonstrations—including the ability to attract participants in the first place—as a paramount concern (Interview, 15 December 2017). Indeed, while the officials we spoke to were well aware of the political challenges associated with expanding successful demonstration projects—especially when congressional approval was required—their immediate focus was on ensuring that demonstrations met internal agency criteria for programme integrity as well as fiscal criteria set by the OMB (Interview, 12 January 2018). Hence even if demonstration projects can be intentionally designed with politics in mind (e.g., by securing buy-in from key stakeholders), individuals with responsibility for implementation may lack the resources, knowledge, or authority to do so.

The findings carry an important implication for related literatures on institutional change and policy experimentation, and suggest that there is great potential in expanding on the link between these literatures. Theories of gradual institutional change hold that bureaucratic actors with a high level of discretion in the implementation or enforcement of legislation can instigate 'conversion', or the 'changed enactment of existing rules due to their strategic redeployment' (Mahoney and Thelen [53], p. 16). Yet our evidence suggests that discretion is a necessary but insufficient condition for stimulating policy change (Shpaizman [80]). Even when Medicare administrators had the discretion to carry out demonstration projects, they could not ensure that these projects would scale up into national reform. Rather, their success at converting Medicare from a fee-for-service programme and towards more 'value-based' payment principles depended on the costs of scaling up demonstration projects into national policy, the mobilization of supporters and opponents of these reforms, as well as the alignment of the project's execution with political time horizons.

Understanding why discretion is insufficient as an explanation of institutional conversion requires a better understanding of the politics of policy experimentation. Even when bureaucrats possess experimental authority, they are not immune to the politics and pathologies of policy learning (Dunlop and Radaelli [26]). Experiments that produce strong evidence from a statistical standpoint may be designed in ways that galvanize entrenched opponents and elicit scepticism from risk-averse policy-

makers. Alternatively, experiments might be designed with input from potential supporters, integrated with a pre-existing organizational infrastructure, or timed to coincide with windows of political opportunity (Stoker and John [83]). What matters for conversion, in short, is not whether policy-makers have discretion to experiment, but how they use that discretion.

While this study has focused on how demonstrations affect policy change at the level of individual policies or programmes, we have not considered how they might contribute to broader change in policy paradigms. Existing literature provides mixed evidence on this front. For example, Steven M. Teles and Timothy S. Prinz (2001) suggest that state-level welfare-to-work demonstrations played an important role in the evolution of conservatives' entitlement reform agenda. Even so, the ambiguity of evidence from demonstration projects and their inability to experiment with policy changes that involve broader economic or structural realities suggest that their role in paradigmatic change will be a limited one (Brodkin and Kaufman [10]). In any case, future research on demonstration projects should attend to the magnitude of policy changes they are able to initiate.

As the increasingly partisan environment of American politics makes even the most incremental processes of change a near Herculean task, it is essential that we gain a better understanding of the alternative pathways to improving public policies. Experimental institutions present one such alternative pathway. This study identifies the political importance of a demonstration's design and its effect on interest group mobilization and the costs of transitioning from experimental setting to national reform. Here we have limited our examination to a small, though emblematic, sample of Medicare demonstrations. To develop a more complete understanding of the politics of experimental institutions, more work is required not only within Medicare, but also comparatively across policy areas, levels of government, and national contexts.

## ACKNOWLEDGEMENTS

The authors wish to thank Adam Sheingate, Colin Moore, and the two anonymous reviewers at *Public Administration* for insightful criticism of earlier drafts of this manuscript. This article also benefited from feedback at the 2016 meeting of the American Political Science Association and the 2017 meeting of the Western Political Science Association.

GRAPH: Appendix: Key Evaluation Documents Reviewed, by Case

## REFERENCES

- Ackermann, R. T. (2017). From programs to policy and back again: The push and pull of realizing type 2 diabetes prevention on a national scale. *Diabetes Care*, 40, 1298 – 1301.
- American Diabetes Association (2016). National Diabetes Prevention Program named the first preventive health initiative eligible for Medicare coverage via CMMI expansion. Available at: <http://www.diabetes.org/newsroom/press-releases/2016/national-dpp-named-first-preventive-health-initiative-eligible-for-medicare.html>.
- American Medical Association (2016a). Diabetes prevention. Available at: <http://0-www.ama-assn.org.libus.csd.mu.edu/ama/pub/advocacy/topics/diabetes-prevention.page>

- American Medical Association (2016b). 'Groundbreaking effort' to prevent diabetes announced. Available at: <http://0-www.ama-assn.org.libus.csd.mu.edu/ama/ama-wire/post/groundbreaking-effort-prevent-diabetes-announced>
- Ball, R. M. (1973). Social security amendments of 1972: Summary and legislative history. *Social Security Bulletin*, 36, 3 – 25.
- Baumgartner, F. R., Berry J. M., Hojnacki, M., Kimball, D. C., & Leech, B. L. (2009). *Lobbying and policy change: Who wins, who loses, and why*. Chicago, IL : University of Chicago Press.
- Béland, D. (2006). The politics of social learning: Finance, institutions, and pension reform in the United States and Canada. *Governance*, 19, 559 – 583.
- Béland, D., & Howlett, M. (2016). How solutions chase problems: Instrument constituencies in the policy process. *Governance*, 29, 393 – 409.
- Blyth, M. (2013). Paradigms and paradox: The politics of economic ideas in two moments of crisis. *Governance*, 26, 197 – 215.
- Brodkin, E. Z., & Kaufman, A. (2000). Policy experiments and poverty politics. *Social Service Review*, 74, 507 – 532.
- Campbell, J. L., & Pedersen, O. K. (2014). *The national origins of policy ideas: Knowledge regimes in the United States, France, Germany, and Denmark*. Princeton, NJ : Princeton University Press.
- Carey, M. A. (2016). Medicare proposes expansion of counseling program for people at risk of diabetes. *Kaiser Health News*, 23 March. Available at: <http://khn.org/news/medicare-proposes-expansion-of-counseling-program-for-people-at-risk-of-diabetes/>
- Carpenter, D. P. (2001). *The forging of bureaucratic autonomy: Reputations, networks, and policy innovation in executive agencies, 1862–1928*. Princeton, NJ : Princeton University Press.
- Carroll, A. (2016). The quiet research that led to a resounding success in diabetes prevention. *New York Times*, 30 March. Available at: <http://0-www.nytimes.com.libus.csd.mu.edu/2016/03/31/upshot/the-unsung-success-of-a-diabetes-prevention-program.html?%5fr=1>
- Carstensen, M. B., & Matthijs, M. (2018). Of paradigms and power: British economic policy making since Thatcher. *Governance*, 31, 431 – 447.
- Cassidy, A. (2008). The fundamentals of Medicare demonstrations. National Health Policy Forum Background Paper No. 63. Available at: <http://www.nhpf.org/library/background-papers/BP63%5fMedicareDemos%5f07-22-08.pdf>
- CMS Office of the Actuary (2016). Certification of Medicare Diabetes Program. Available at: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/ActuarialStudies/Downloads/Diabetes-Prevention-Certification-2016-03-14.pdf>
- Coyle, D. J., & Wildavsky, A. (1986). Social experimentation in the face of formidable fables. In A. Munnell (Ed.), *Lessons from income maintenance experiments* (pp. 167 – 184). Boston, MA : Federal Reserve Bank of Boston.
- Cromwell, J., Dayhoff, D., McCall N., Subramanian, S., Freitas, R., & Hart R. (1998). *Medicare Participating Heart Bypass Center demonstration: Final report*. Waltham, MA : Health Economics Research.
- Davis, M., & Burner, S. T. (1995). Three decades of Medicare: What the numbers tell us. *Health Affairs*, 14, 231 – 243.

- Diabetes Advocacy Alliance (2016). Letter to the Bipartisan Chronic Care Working Group. 26 January. Available at: <http://www.finance.senate.gov/imo/media/doc/Diabetes%20Advocacy%20Alliance%20comments%20to%20SFC%20CCWG%20012616.pdf>
- Diabetes Prevention Program Research Group (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *New England Journal of Medicine*, 346, 393 – 403.
- Dunlop, C. A. (2016). Contestation and contingency in advisory governance. In M. Bevir & R. A. W. Rhodes (Eds.), *Rethinking governance: Rules, rationalities and resistance* (pp. 162 – 178). London : Routledge.
- Dunlop, C. A. (2017). Pathologies of policy learning: What are they and how do they contribute to policy failure? *Policy and Politics*, 45, 19 – 37.
- Dunlop, C. A., & Radaelli, C. M. (2013). Systematising policy learning: From monolith to dimensions. *Political Studies*, 61, 599 – 619.
- Dunlop, C. A., & Radaelli, C. M. (2018). The lessons of policy learning: Types, triggers, hindrances and pathologies. *Policy and Politics*, 46, 255 – 272.
- Elmore, R. (1986). A political scientist's view of the income maintenance experiments. In A. Munnell (Ed.), *Lessons from income maintenance experiments* (pp. 206 – 213). Boston, MA : Federal Reserve Bank of Boston.
- Evans, M. (2014). Medicare ACOs can learn lessons from earlier demonstration projects. *Modern Healthcare*, 26 August. Available at: <http://www.modernhealthcare.com/article/20140826/NEWS/308269936>
- Fisher, E. S. (2009). The implications of regional and provider-specific variations in Medicare spending for Medicare payment reform. Testimony before the Committee on Ways and Means, US House of Representatives, 1 April.
- Flanigan, J. (1990). Health care suffers unkindest cut of all. *Los Angeles Times*, 3 October, D1.
- George, A., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, MA : The MIT Press.
- Gold, M. (2010). *Accountable care organizations: Will they deliver?* Princeton, NJ : Mathematica Policy Research.
- Government Accountability Office (2018). *Medicaid demonstrations: Evaluations yielded limited results, underscoring need for changes to federal policies and procedures*. GAO-18-220. Washington, DC : Government Printing Office.
- Granger, R. C., & Maynard, R. (2015). Unlocking the potential of the 'what works' approach to policymaking and practice: Improving impact evaluations. *American Journal of Evaluation*, 36, 558 – 569.
- Grassley, C. (2016). Grassley welcomes proposed expansion of Diabetes Prevention Program. Available at: <http://www.grassley.senate.gov/news/news-releases/grassley-welcomes-proposed-expansion-medicare-diabetes-prevention-program>
- Greenwald, L. (2011). Converting successful Medicare demonstrations into national programs. In J. Cronwell, M. Trisolini, G. Pope, J. B. Mitchell, & L. M. Greenwald (Eds.), *Pay for performance in health care: Methods and approaches* (pp. 315 – 340). Research Triangle Park, NC : RTI Press.
- Guterman, S., Davis, K., Stremikis, K., & Drake, H. (2010). Innovation in Medicare and Medicaid will be central to health reform's success. *Health Affairs*, 29, 1188 – 1193.

- Hacker, J. S. (2004). Privatizing risk without privatizing the welfare state: The hidden politics of social policy retrenchment in the United States. *American Political Science Review*, 98, 243 – 260.
- Hacker, J. S., Pierson, P., & Thelen, K. (2015). Drift and conversion: Hidden faces of institutional change. In J. Mahoney & K. Thelen (Eds.), *Advances in comparative-historical analysis* (pp. 180 – 208). New York : Cambridge University Press.
- Hansen, J. C. (2009). Statement before the Energy and Commerce Health Subcommittee. US House of Representatives, 23 June.
- Herdman, R. C., Wagner, J. L., Wolfe, L. Gluck, M. E., & Brogan, T. W. (1993). Health technology and health reform. *Cancer Investigation*, 11, 337 – 344.
- Immergut, E. M. (1992). *Health politics: Interests and institutions in Western Europe*. New York : Cambridge University Press.
- Institute of Medicine (IOM) (1999). *To err is human: Building a safer health system*. Washington, DC : National Academies Press.
- Jacobson, N., & Jones, T. (2001). BIPA: The Medicare, Medicaid, and SCHIP Benefits Improvement and Protection Act of 2000. *Policy, Politics, and Nursing Practice*, 2, 228 – 232.
- Jowell, R. (2003). *Trying it out: The role of 'pilots' in policy-making*. London : Cabinet Office.
- Kahn, C., & Kuttner, H. (1999). Budget bills and Medicare policy: The politics of the BBA. *Health Affairs*, 18, 37 – 47.
- Kautter, J., Pope, G., Leung, M., Trisolini, M., Adamache, W., Smith, K., ... & Schwartz, M. (2010). *Evaluation of the Medicare Physician Group Practice demonstration: Final report*. Chicago, IL : RTI International.
- Kautter, J., Pope, G., Trisolini, M., & Grund, S. (2007). Medicare Physician Group Practice demonstration design: Quality and efficiency pay-for-performance. *Health Care Financing Review*, 29, 15 – 29.
- Khatana, S. A., Albright, A. L., & Sanghavi, D. M. (2016). JAMA forum: Expanding Medicare access to the Diabetes Prevention Program. 23 March. Available at: <https://newsatjama.jama.com/2016/03/23/jama-forum-expanding-medicare-access-to-the-diabetes-prevention-program/>
- Kingdon, J. (2011). *Agendas, alternatives, and public policies* (2nd ed.). New York : Pearson.
- Kirsch, R. (2009). Letter to George Miller, Charles Rangel, Henry Waxman, Robert Andrews, Pete Stark, and Frank Pallone. Re: Comments on House Tri-Committee Draft Health Care Reform Legislation. 2 July.
- Levy, J. M., & Borowitz, M. (1992). The Medicare fee schedule unveiled: An account of physician payment reform. *Health Policy*, 22, 263 – 286.
- Mahoney, J., & Thelen, K. (Eds.) (2010). *Explaining institutional change: Ambiguity, agency, and power*. New York : Cambridge University Press.
- Mahoney, J., & Thelen, K. (2015). Comparative-historical analysis in contemporary political science. In J. Mahoney & K. Thelen (Eds.), *Advances in comparative-historical analysis* (pp. 3 – 38). New York : Cambridge University Press.
- Marino, M., Bailey, S. R., Gold R., Hoopes, M. J., O'Malley, J. P., Huguet, N., ... & DeVoe, J. E. (2016). Receipt of preventive services after Oregon's randomized Medicaid experiment. *American Journal of Preventive Medicine*, 50, 161 – 170.
- Marwick, C. (1993). 'Global package' saves money but stirs criticism. *JAMA*, 270, 927.

- Matthijs, M., & Blyth, M. (2018). When is it rational to learn the wrong lessons? Technocratic authority, social learning, and euro fragility. *Perspectives on Politics*, 16, 110 – 126.
- McKay, A., & Yackee, S. W. (2007). Interest group competition on federal agency rules. *American Politics Research*, 35, 336 – 357.
- Mechanic, D. (2001). The managed care backlash: Perceptions and rhetoric in health care policy and the potential for health care reform. *Milbank Quarterly*, 79, 35 – 54.
- Medicare Payment Advisory Commission (MedPAC) (2009). Report to the Congress: Improving incentives in the Medicare program. Washington, DC : MedPAC.
- Millenson, M. (1988). Heart bypass surgery varies: Report shows wide range in costs and hospital stays. *Chicago Tribune*, 4 August, A6.
- Miller, M. (2008). Statement before the Committee on Finance, US Senate. 16 September.
- Nathan, R. (2000). *Social science in government: The role of policy researchers*. Albany, NY : Rockefeller Institute Press.
- Oakley, A. (1998). Public policy experimentation: Lessons from America. *Policy Studies*, 19, 93 – 114.
- Oberlander, J. (2003). *The political life of Medicare*. Chicago, IL : University of Chicago Press.
- O'Connor, A. (2009). *Poverty knowledge: Social science, social policy, and the poor in twentieth-century US history*. Princeton, NJ : Princeton University Press.
- Orren, K., & Skowronek, S. (2017). *The policy state: An American predicament*. Cambridge, MA : Harvard University Press.
- Patashnik, E. M., & Zelizer, J. E. (2013). The struggle to remake politics: Liberal reform and the limits of policy feedback in the contemporary American state. *Perspectives on Politics*, 11, 1071 – 1087.
- Pear, R. (1999). Congress rejects idea of using managed care for Medicare. *New York Times*, 22 October. Available at: <http://www.nytimes.com/1999/10/22/us/congress-rejects-idea-of-using-managed-care-for-medicare.html>.
- Peck, J., & Theodore, N. (2015). *Fast policy: Experimental statecraft at the thresholds of neoliberalism*. Minneapolis, MN : University of Minnesota Press.
- Peck, L. R. (2016). On the 'how' of social experiments: Analytic strategies for getting inside the black box. *New Directions for Evaluation*, 152, 85 – 96.
- Peterson, M. (2015). Third rail of politics: The rise and fall of Medicare's untouchability. In A. B. Cohen, D. C. Colby, K. A. Wailoo, & J. E. Zelizer (Eds.), *Medicare and Medicaid at 50: America's entitlement programs in the age of the Affordable Care Act* (pp. 147 – 168). New York : Oxford University Press.
- Ratner, R. E., Stack, S. J., Longjohn, M., & Payne, M. (2016). Fighting a national epidemic by investing in prevention. *Huffington Post*, 26 April. Available at: <http://www.huffingtonpost.com/steven-j-stack/fighting-a-national-epide%5fb%5f9780050.html>
- Rich, S. (1987). Bypass surgery safer, cheaper in big hospitals. *The Washington Post*, 20 August, A15.
- Rosenbaum, S. (1992). Mothers and children last: The Oregon Medicaid experiment. *American Journal of Law and Medicine*, 18, 97 – 126.
- Rosenblatt, R. A. (1990). Medicare cutting costs: 4 to 6 hospitals to be chosen for discount bypass surgery. *Los Angeles Times*, 27 June, A5.
- Saldin, R. (2017). *When bad policy makes good politics: Running the numbers on health reform*. New York : Oxford University Press.

- Schmidt, V. A. (2016). The roots of neo-liberal resilience: Explaining continuity and change in background ideas in Europe's political economy. *British Journal of Politics and International Relations*, 18, 318 – 334.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA : Houghton Mifflin.
- Shpaizman, I. (2014). Ideas and institutional conversion through layering: The case of Israeli immigration policy. *Public Administration*, 92, 1038 – 1053.
- Steinmo, S., & Watts, J. (1995). It's the institutions, stupid! Why comprehensive national health insurance always fails in America. *Journal of Health Politics, Policy and Law*, 20, 329 – 372.
- Stoker, G. (2010). Translating experiments into policy. *Annals of the American Academy of Political and Social Science*, 628, 47 – 58.
- Stoker, G., & John, P. (2009). Design experiments: Engaging policy makers in the search for evidence about what works. *Political Studies*, 57, 356 – 373.
- Stout, H. (1991). Medicare starts program to curb the costs of heart bypass surgery. *Wall Street Journal*, 31 January, B4.
- Tanenbaum, S. J. (2009). Pay for performance in Medicare: Evidentiary irony and the politics of value. *Journal of Health Politics, Policy and Law*, 34, 717 – 746.
- Teles, S. M., & Prinz, T. S. (2001). The politics of rights retraction: Welfare reform from entitlement to block grant. In M. Levin, M. K. Landy, & M. Shapiro (Eds.), *Seeking the center: Politics and policymaking at the new century* (pp. 215 – 238). Washington, DC : Georgetown University Press.
- Thompson, F. J. (2012). *Medicaid politics: Federalism, policy durability, and health reform*. Washington, DC : Georgetown University Press.
- Tompkins, C. P., Wallack, S. S., Bhalotra, S., Chilingirian, J. A., Glavin, M. P. V., Ritter, G. A., & Hodgkin, D. (1996). Bringing managed care incentives to Medicare's fee-for-service sector. *Health Care Financing Review*, 17, 43 – 63.
- US Congress. House. Committee on the Budget (2010). *Report of the Committee on the Budget, House of Representatives to Accompany H.R. 4872*. 111th Cong., 2nd sess., 17 March.
- US Congress. House. Committee on Energy and Commerce, Subcommittee on Health (2009). *Comprehensive health reform: Discussion draft: Hearing before the Subcommittee on Health*. 111th Cong., 1st sess., 23–25 June.
- US Congress. House. Committee on Ways and Means (1997). *Recommendations regarding Medicare hospital and physician payment policies: Hearing before the Subcommittee on Health of the Committee on Ways and Means, House of Representatives*. 105th Congress, 1st sess., 20 March.
- US Congress. House. Committee on Ways and Means (2009). *Health reform in the 21st century: Reforming the health care delivery system: Hearing before the Committee on Ways and Means*. 111th Cong., 1st sess., 1 April.
- US Congress. Senate. Finance Committee (2015). *A pathway to improving care for Medicare patients with chronic conditions*. 114th Congress. 1st sess., 14 May.
- Wallack, S. S., & Tompkins, C. P. (2003). Realigning incentives in fee-for-service Medicare. *Health Affairs*, 22, 59 – 70.

- Wallis, J. J. (1998). The political economy of New Deal spending revisited, again: With and without Nevada. *Explorations in Economic History*, 35, 140 – 170.
- Weight Watchers (2016). Letter to the Bipartisan Chronic Care Working Group. 26 January. Available at: <http://www.finance.senate.gov/imo/media/doc/Weight%20Watchers%20Comment%20Letter%20on%20%20SFCCWG%20Policy%20Porposal.pdf>
- Wilensky, G. R. (2009). Slow spending, improve quality in Medicare: Commentary on the Health Care Opinion Leaders Survey. Commonwealth Fund blog, 2 November. Available at: <https://www.commonwealthfund.org/publications/publication/2009/nov/slow-spending-improve-quality-medicare-commentary-health-care>.
- Williamson, T. (2009). Statement before the Committee on Energy and Commerce, Subcommittee on Health, US House of Representatives, 25 June.
- Yackee, J. W., & Yackee, S. W. (2006). A bias towards business? Assessing interest group influence on the US bureaucracy. *Journal of Politics*, 68, 128 – 139.
- Zarabozo, C. (2000). Milestones in Medicare managed care. *Health Care Financing Review*, 22, 61 – 67.