Received: 28 March 2021 | Revised: 28 September 2021 | Accepted: 7 October 2021

DOI: 10.1111/padm.12790

ORIGINAL ARTICLE

public administration WILEY

"Success" in policy piloting: Process, programs, and politics

Kath Checkland[1] | Jonathan Hammond[1] | Anna Coleman[1] | Julie Macinnes[2] | Rasa Mikelyte[2] | Sarah Croke[1] | Jenny Billings[2] | Simon Bailey[2] | Pauline Allen[3]

[1]School of Health Sciences, University of Manchester, Manchester, UK

[2]Centre for Health Services Studies, University of Kent, Canterbury, UK

[3]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

Correspondence
Kath Checkland, School of Health Sciences, University of Manchester, Manchester, UK.
Email: katherine.checkland@manchester.ac.uk

Funding information
NIHR Policy Research Programme, Grant/Award Number: PR-R16-0516-22001

Abstract

Research has demonstrated that pilots contain multiple shifting purposes, not all of which relate to simple policy testing or refinement. Judging the success of policy pilots is therefore complex, requiring more than a simple judgment against declared goals. Marsh and McConnell provide a framework against which policy success can be judged, distinguishing program success from process and political success. We adapt Boven's modification of this framework and apply it to policy pilots, arguing that pilot process, outcomes and longer-term effects can all be judged in both program and political terms. We test this new framework in a pilot program in the English National Health Service, the Vanguard program, showing how consideration of these different aspects of success sheds light on the program and its aftermath. We consider the implications of the framework for the comprehensive and multifaceted evaluation of policy pilots.

This paper is based on independent research commissioned and funded by the NIHR Policy Research Programme "National evaluation of the Vanguard New Care Models Programme," PR-R16-0516-22001. The views expressed in the publication are those of the author(s) and not necessarily those of the Policy Research Programme, NIHR, or the Department of Health and Social Care.

1 | INTRODUCTION

The piloting of policy initiatives prior to wider rollout (sometimes termed "policy experiments") is increasingly popular in many jurisdictions (Heilmann, 2008; Nair & Howlett, 2016; Tassey, 2013), in part because of an implicit (and sometimes explicit) association with apparently rational and depoliticized policy making (Brodkin & Kaufman, 2000; Martin & Sanderson, 1999). A UK Cabinet Office report in 2003 recommended piloting as the default approach to policy making (Jowell, 2003), identifying a rational process by which pilots should be carefully evaluated prior to decisions about wider rollout. Rogers-Dillon (2004, p. 24) considers this vogue for policy piloting and identifies a yearning for a "cool, pristine world of policy," technical, efficient, and removed from the "messy world of politics."

However, the reality is considerably more complex than this ideal implies. Exploring the operation of policy pilots in the English National Health Service, Ettelt et al. (2014) suggest that, alongside rational testing of policy ideas, pilots are also used to expedite implementation, with the "success" of a pilot in meeting its goals less important than its ability to catalyze implementation. Moreover, Rogers-Dillon (2004) highlights the potential party political or ideological effects of pilots, arguing that the mere existence of "workfare" pilots (i.e., programs linking welfare entitlements to engagement with work) in some US states in the 1990s shifted public and political opinion, rendering mainstream approaches to welfare which had previously been unthinkable. These effects were not linked to pilot outcomes, nor conduct, but rather the fact that their very existence overcame deeply embedded prior assumptions about fairness and equity in welfare programs. Nair and Howlett (2016) situate pilots as "framing or projecting the future" (p. 1), establishing meaning associated with policies and expressing and enacting power relationships, while Bailey et al. (2017) show local pilots providing a political narrative which enhanced an existing national policy agenda.

How then, is the success of any given policy pilot to be judged? A straightforward experimentation approach implies the rational pre-specification of desired outcomes, followed by explicit judgment of their achievement, but the more messy and emergent process described by Ettelt et al. (2014), Bailey et al. (2017), and Rogers-Dillon (2004) is not so easily judged. The question of general policy success has been addressed by a number of authors, led by Marsh and McConnell (Marsh & McConnell, 2010; McConnell, 2010) who identify three dimensions of policy success: process success, the successful making of policy, including passing legislation or creating a supportive coalition; program success, encompassing successful policy implementation and achievement of desired goals; and political success, the potential for policies to enhance a government's reputation or advance its ideology. Bovens (2010) modifies this framework, arguing that both policy making and policy implementation can be considered in program and political terms. In this paper we contend that such frameworks require further modification to address the nature of pilots as policy projects limited in time and occurring in specific places (Bailey et al., 2019), embodying complex and not necessarily straightforward purposes. We offer a modified framework to consider policy pilot success, and test it using an example of policy piloting in England.

Our policy example is the Vanguard New Care Models program in the NHS England (2014a). These well-funded and supported pilots were intended to derive and test more integrated ways of providing health and care services (NHS England, 2015b). Officially designated as a "successful" program which has demonstrated better ways of designing services (NHS England, 2019b), we draw upon an evaluation of the program to consider this claim. Using our modified framework, we show that, while the program met some outcome goals, longer-term local impact and further rollout have been limited. Moreover, we show how early political claims about success may have impeded the work required to facilitate either wider rollout or systematic learning from the pilots. This latter finding demonstrates the importance of including a political dimension in evaluating pilot success. We conclude by highlighting the value of our framework in supporting multifaceted and comprehensive evaluations of policy pilots, which in turn may enable better understanding of later policy trajectories.

1.1 | Policy success and failure

McConnell (2010) argues for a more nuanced approach to policy success than one focusing upon the achievement of predefined goals. He suggests that the success or failure of any given policy will be plural and contested, and that

the policy sciences: "lack an over-arching heuristic framework which would allow analysts to approach the multiple outcomes of policies in ways that move beyond the often crude, binary rhetoric of success and failure" (McConnell, 2010, p. 346).

He goes on to define such a framework, identifying three dimensions of policy success. *Process success* refers to the policy making process. McConnell (2010) suggests a successful policy process is one in which a "Government does what it sets out to do and opposition is virtually non-existent and support near universal" (McConnell, 2010, p. 352). Dimensions of success include the establishment of a stable coalition behind a policy, and lack of significant opposition. *Program success* encompasses successful implementation and the achievement of desired goals, including benefits for targeted populations. Finally, *political success* refers to the political benefits accruing out of a particular policy. These might include sustaining an electoral coalition, enhancing electoral prospects, or silencing opposition.

Under each heading McConnell (2010) identifies a spectrum, from "success" with criteria fully met, through to "failure" when none of the criteria are met. In between—the "gray area"—success may be partial, allowing more sophisticated judgments as to policy impacts (Marsh & McConnell, 2010).

However, the complex formulation of dimensions of success across the three domains has been criticized. In a commentary accompanying the paper by Marsh and McConnell (2010), Bovens (2010) argues that the authors make a category mistake in separating policy process, programs and politics. He argues that policy process—that is, the processes surrounding policy enactment—occupies a different analytical level to program and political outcomes and can itself be evaluated both programmatically AND politically. Thus, for Bovens, "policy process success" can be both programmatic—for example, when a piece of legislation is successfully maneuvered through parliament—AND political—when the passage of such legislation enhances a government's political capital. Bovens (2010) goes on to argue for a categorization of policy success or failure across two dimensions: process and outcomes; and program and politics (see Table 1).

This approach thus distinguishes between political and program success (which Bovens calls the "focus" or perspective from which success is to be judged) and applies this to both policy formation and policy outcomes (the "locus" or object of the assessment). This has the advantage of separating the political benefits or disbenefits of *policy making* from those associated with *policy implementation*.

More recently, focus has expanded to consider the temporal nature of judgments about policy success: at what point is it appropriate to judge a policy? In a recent book, t'Hart (2019) explores policy successes, adding the question of policy sustainability to Marsh and McConnell's (2010) framework for judging success. He argues:

> A policy is a complete success to the extent that (a) it demonstrably creates widely valued social outcomes; through (b) design, decision-making, and delivery processes that enhance both its problem-solving capacity and its political legitimacy; and (c) sustains this performance for a considerable period of time, even in the face of changing circumstances. (t'Hart, 2019, p. 5)

Thus, he suggests that only policies which endure and deliver ongoing public value can be truly designated as "successful." However, it remains unclear exactly what "a considerable period" might be. Indeed, it could be argued

TABLE 1 Dimensions of policy success

	Program success	Political success
Policy process	Policy developed as planned, legislation successfully passed	Passing the legislation or developing the policy enhanced the government's reputation or electoral prospects
Policy outcome	Policy implemented as planned, policy outcomes achieved	The implementation or outcome of the policy enhanced the government's reputation

Source: Adapted from Bovens (2010).

that what constitutes a meaningful endurance of particular policies will depend upon such things as changes of government, with endurance beyond the hegemony of a particular political party potentially indicative of sustained success even if the absolute timescale remains short. Moreover, "endurance" may, as demonstrated by Rogers-Dillon (2004), be less to do with programmatic endurance of a particular policy initiative and more to do with a long-term shift in how society views a particular issue. The judgment of the extent of policy success becomes yet more complex when considering what Newman (2014) calls the "distributional" question, arguing that McConnell and Marsh fail to take account of the differential impact of policies on different sectors of society. McConnell et al. (2020) take this further, providing a framework for considering not only differential *societal* impacts, but also the impact on actors at each level of the process: policy making; policy implementation and enactment; and politics. Thus, for example, a policy might benefit one political actor over another, enhancing their reputation and providing further opportunities for influence, while at the same time providing material benefits to a particular sector of society.

Thus, a complex and multifaceted set of frameworks for judging policy success emerges, defining success from multiple perspectives, across time and through varied lenses, differentiating between material outcomes and those of a more political nature. Importantly, these approaches draw attention to the fact that desired outcomes might not be fully declared, and, as highlighted by such policy analysis approaches such as the advocacy-coalition framework (Sabatier, 2006; Sabatier & Weible, 2014) and Kingdon's (1995) streams and windows, will usually entail agendas beyond the desire to provide public value.

1.2 | Success and policy pilots: An adapted framework

How then, should we consider judging the success or otherwise of policy pilots? Pilots differ from full policy implementation in that they are limited in both time and space (Bailey et al., 2019). Moreover, they embody the ostensibly rational purpose of testing potential policy solutions; they therefore rhetorically at least embody some uncertainty as to their value. Importantly, the temporal dimension of policy success expounded by t'Hart (2019) must be considered, with the longer-term rollout or spread of piloted policies an important element to be judged.

Much literature on policy piloting situates the use of pilots within the assumptions of evidence-based policy. as small-scale experiments testing a policy prior to wider rollout (Burch & Wood, 1983; Jowell, 2003). In practice these assumptions are challenged by the political constructions and uses of knowledge and evidence within pilots (Martin & Sanderson, 1999; Sanderson, 2002). We expand upon these concerns to consider the performative effect of politics upon the purposes and outcomes of pilots.

Moving beyond the notion of experimentation, Harrison and Wood (1999), show how "manipulated emergence" arises out of incentivized early adoption of loosely defined "bright ideas," which is argued to be more effective than conventional top-down implementation. This suggests a more "generative" understanding of experimentation (Ansell & Bartenberger, 2016), alongside other implicit purposes, such as exemplification. Ettelt et al. (2014) extend this, suggesting that pilot programs may be driven by a variety of purposes (experimentation, demonstration, early adoption, and learning), only some of which might be explicit, and which might shift and intersect during the program. In this context, "success" is not a simple concept, and requires an analytical approach which is attuned to the different political "levels" which piloting traverses, as well as the temporal dimension of success implicit in moves from temporary pilots to enduring organizational arrangements (Bailey et al., 2019).

Taking these issues into account, we draw upon Bovens' (2010) modification of Marsh and McConnell's (2010) framework. Agreeing with Bovens that both policy process and policy outcomes can and should be judged in both program and political terms, we argue that, in keeping with Ettelt et al.'s analysis, the longer-term rollout or termination of pilot programs should also be considered across these two dimensions (see Table 2).

Thus, we suggest that, in addition to considering the program and political effects of the design and implementation of policy pilots, a longer-term view of the eventual impact of pilots should also be considered. We have termed

this "pilot effects." By this we mean the effect of the pilot beyond the immediate judgment as to whether ostensible goals have been met, and beyond the term of the pilot. In program terms, taking the ostensibly rational view, pilots should be judged according to whether or not apparently beneficial effects are implemented more widely, or, if assessment demonstrates no obvious benefits, the pilots are rationally modified or terminated. More widely, pilot program effects might also be to influence the shape or direction of future policy. The effect of the pilot from a political perspective, however, can be more complex to assess, as it may range from party political advantage through to less obvious accrual of power or advantage to one or more actors within the system. For example, in Rogers-Dillon's (2004) example of workfare pilots in the United States, political advantage accrued to the party in power, shifting public opinion to allow further changes to the welfare system without attracting electoral disadvantage. Alternatively, a pilot program might empower a particular nongovernmental body, enhancing their influence by association with an ostensibly successful pilot. Taking the view, with Lasswell (1936 [2018]), that politics relates to the question of "who gets what, when, how?," we argue that judgment of the political success or otherwise of pilots should consider how policies influence the distribution of power or resources in a political system alongside party political advantage. This political dimension is particularly important in considering pilot effects, given their multiple, shifting and potentially undeclared purposes (Ettelt et al., 2014). However, we acknowledge that judging political success will be multifaceted and complex. In this paper, we explore the effect of pilots on the distribution of power and resources; we do not explore in depth more diffuse questions of complex political goals or hidden agendas. We return to this question in our discussion.

Finally, in considering McConnell et al.'s (2020) distributional question, we acknowledge that any assessment of a policy pilot's success will be from a particular perspective. A pilot which acts to empower one actor will often disempower another, and it is therefore important that *the perspective from which success is being judged* should be declared in operationalizing the framework.

Population of the "cells" in the framework requires evaluative work utilizing multiple methods, from quantitative analysis of before-and-after outcomes to qualitative analysis examining political speeches and documents. In order to explore the utility of this framework in judging different aspects of "success" as applied to policy pilots, we here apply it to an English national policy pilot scheme, the Vanguard New Care Models program, combining findings from a variety of evaluative methods and approaches to produce an overall assessment of "success."

1.3 | Research context: The NHS in England and the Five Year Forward View

NHS England was created in 2012 as an Arm's length Body, responsible for the delivery of health services to the population under a "mandate" from the Department of Health and Social Care (Hammond et al., 2018). In 2014 NHS England published a policy document, the Five Year Forward View (NHS England, 2014a), setting out the challenges

TABLE 2 Dimensions of policy pilot success

	Program success	Political success
Pilot process	Did the piloting program happen—that is, was it developed and implemented?	Did pilot initiation have any positive political consequences, for whom?
Pilot outcomes	Did the piloting program meet its ostensible goals?	Did its eventual outcome have any positive political consequences, for whom?
Pilot effects	Was the program locally sustained and/or more generally rolled out? OR was it rationally modified or discontinued? Was future policy altered as a result?	Did rollout or discontinuation have positive political consequences, for whom?

facing the NHS, including demographic pressures, and shrinking budgets following the global financial crash. The solution offered was increasing integration between different types of providers, "dissolving traditional boundaries" and "learning fast from the best examples" (p. 16). The document proposed the creation of pilots—known as Vanguards—to test out new ways of providing services. A number of new service models were suggested, eventually consolidated into five different types of Vanguard (Table 3). Local areas were invited to apply, and, following a selection process, 50 sites were chosen and provided with additional funding as set out in Table 3. The substance of the service delivery models was left for the sites to determine.

An extensive support program was established, alongside a formal evaluation program. The findings presented here draw upon an independent national evaluation of the program, commissioned and funded by the National Institute for Health Research Policy Research Programme (Checkland et al., 2019; Checkland et al., 2021). This evaluation focused upon the three Vanguard types which addressed integration between hospital, community, and social care services (multispecialty community providers [MCPs], primary and acute care systems [PACS], and enhanced health in care homes [ECHs]).

Crucially, the Five Year Forward View argued that "one size will not fit all" (NHS England, 2014a, p. 9), with diversity of local solutions encouraged. The well-resourced support package, extensive program of continuous evaluation and expectation of local determination makes explicit a rationale of "generative" rather than "controlled"

TABLE 3 Types of Vanguard

Vanguard type	Date	Number	Description	Funding between 2015–2016 and 2017–2018 (£ million)
Primary and acute care systems (PACS)	March 2015	9	Joining up GP, hospital, community, and mental health services to improve the physical, mental, social health, and wellbeing of the local population. Population-based care model based on the GP registered list	103
Multispecialty community providers (MCPs)	March 2015	14	Moving specialist care out of hospitals into the community. Working to develop population-based health and social care. Population-based care model based on the GP registered list	124
Enhanced health in care homes (ECH)	March 2015	6	Offering older people better, joined up health, care, and rehabilitation services. Care homes working closely with the NHS, local authorities, the voluntary sector, carers, and families to optimize health of their residents	18
Urgent and emergency care networks (UECs)	July 2015	8	New approaches to improve the coordination of services and reduce pressure on A&E departments	72
Acute care collaboratives (ACCs)	September 2015	13	Linking local hospitals together to improve their clinical and financial viability, reducing variation in care and efficiency	13 (1 year only)

Source: Funding amounts derived from National Audit Office (2018, p. 6).

experimentation (Ansell & Bartenberger, 2016). By this we mean that the design of the program suggested a desire to: "generat[e] and iteratively refin[e] a solution concept (an idea, innovation, design, policy, program, etc.) based on continuous feedback and with the goal of addressing a particular problem" (Ansell & Bartenberger, 2016, p. 68). Controlled experimentation, by contrast, would have initiated clearly delineated and characterized programs, with before-and-after analysis of outcomes.

The program ran for 3 years from 2015. Vanguard pilots received £329 million, with £60 million spent on support and evaluation (Checkland et al., 2019; National Audit Office, 2018). The program is referred to repeatedly in subsequent policy documents as having been "successful" (NHS England, 2017, 2019a, 2019b).

In the rest of this paper, we will explore the program in depth, and consider this "success" using the framework which we have derived from Bovens (2010). A final section considers the value of this approach to exploring the success or otherwise of policy pilots.

2 | METHODS

The paper draws upon a wider evaluation program, the findings of which are reported elsewhere (Checkland et al., 2019; Coleman, Billings, et al., 2020a; Morciano et al., 2020). In this paper, we look across the data collected to answer the questions:

- To what extent can the Vanguard pilot program be judged successful, across which dimensions?
- Does our proposed framework capture relevant aspects of pilot success, and how might it be improved?

We draw upon the findings from three elements of the research: initial qualitative study of program initiation and oversight; qualitative case studies exploring program operation; and an ongoing study of relevant policy documents, including those which use the Vanguard program to make arguments about future policy direction. We contextualize our findings with reference to a quantitative impact analysis, published elsewhere (Morciano et al., 2020).

We first analyzed all policy documents produced by NHSE to support the program (NHS England, 2014a, 2014b, 2015a, 2015b, 2017), focusing upon understanding the espoused program goals. Our analysis (Checkland et al., 2019) suggested that these were:

- To implement integrated care programs in designated Vanguard areas.
- To use Vanguard experiences to design "standard approaches and products" which could be rolled out.
- To monitor performance against "benchmarks" and use this information to guide future investment decisions.

We then interviewed 29 stakeholders, including senior managers from NHS England, regional staff supporting local Vanguards, members of a national oversight group, and representatives of the national regulators, NHS Improvement and the Care Quality Commission. Interviewees were purposely selected to represent the principal groups of stakeholders involved in the program, including senior managers responsible for its initiation, those leading the program, those responsible for day to day running and those responsible for associated evaluative activity. In addition, we used "snowball" sampling to identify key individuals with knowledge about particular aspects of the program, including regulatory representatives and those with an advisory/oversight role. Table 4 details the interviewees.

Interviews were semi-structured, with tailored topic guides for each group of interviewees. The focus was upon their experience of the planning, initiation, and operation of the pilot program, and its outcomes. The program formally commenced in April 2015, and concluded at the end of March 2018. The interviews took place in years 2 and 3. Interviews were transcribed verbatim, and analyzed using the computerized analysis program NVivo.

The second phase of the study took a qualitative case study approach to explore the processes and experiences of participants involved in implementing and operating the Vanguard program (2015–2018) at the local level. We

TABLE 4 Interviewees in phase 1

Interviewee type	Numbers interviewed
NHS England employees (current/past)	19
Advisors	7
Regulators	3

TABLE 5 Interviewees and focus groups phase 2

Respondent type	Numbers interviewed	Numbers in focus groups
NHS employees (current/past)	48	14
Local authority	4	–
Private/community/charity sector	9	1
Public contributor	1	4

selected six case-study sites to study in depth: two MCPs, two PACS and two ECH Vanguards. Between October 2018 and July 2019, we carried out focus groups and interviews with a variety of respondents at six case study sites. Individual interviews were used to elicit individual participants considered reflections about their personal roles and experiences in the program. Focus groups were used as a means of eliciting reflective discussion among groups (Nyumba et al., 2018). This approach was particularly used among groups of senior executives involved in the pilots and with public contributors, asking these contributors to retrospectively reflect upon historical events and decision-making processes. The focus group approach was particularly valuable in eliciting illuminating discussions which allowed us to understand from a variety of perspectives and in more depth how particular decisions came to be made. Interviews were a mix of face-to-face or telephone. Focus groups were conducted face-to-face and facilitated by at least one researcher. A total of 80 respondents participated across the sites, including current and past representatives from Clinical Commissioning Groups (CCG), provider organizations, local authorities, voluntary sector organizations, Vanguard program leads, frontline staff and patient/public contributors. One NHS employee participated in both an interview and a focus group (Table 5).

Focus groups and interviews were recorded and transcribed verbatim, followed by a thematic analysis using a coding schedule based on previous literature and our previous findings (Checkland et al., 2019) using NVivo software.

Our ongoing analysis of policy documents involves systematic capture and analysis of all major policy documents issued by NHS England or the Department of Health and Social Care from 2014. All new documents are read and interrogated for mention of the Vanguard program. The content of relevant extracts is explored to consider: the context in which the Vanguard program is mentioned; any claims made about its success; and the rhetorical uses made of any such claims. Our approach to this process is interpretive, viewing policy documents as pieces of rhetoric, seeking to make an argument (Winton, 2013). Our aim was therefore to understand which particular arguments the performance of the Vanguard program was used to underpin, in which contexts, in order to better understand how the program is being used politically, by whom.

For the purposes of this paper, all sources of data were synthesized and a second order analysis undertaken to consider pilot "success" against our framework.

3 | FINDINGS

In this section, we apply our framework for exploring policy success to the Vanguard program. In keeping with our appreciation of the importance of McConnell et al.'s (2020) distributional question (success for whom?), our

perspective is that of those initiating the pilots, NHS England. We return to the question of other perspectives in our discussion.

3.1 | Pilot process program success: Was the program successfully developed and initiated?

From this viewpoint, the program was a resounding success. The timescale involved was extremely tight: the Five Year Forward View (NHS England, 2014a) proposing the pilot program was published in September 2014, with details about how to apply to join the program announced in December and the first 29 Vanguard sites chosen in March 2015. Pilot initiation followed within months. At the same time a wide-ranging support program was set up (NHS England, 2015b), providing individual "account managers" for each chosen site, alongside workstreams addressing anticipated issues in integrating care across organizational and sector boundaries.

By September 2015 (only 6 months after pilot initiation), all 50 sites were in operation across all five "models," a designated National lead, Vanguard "model" leads, support stream leads and strategic account managers appointed, funding distributed to sites and the support program in operation. In program terms, therefore, the pilot process was extremely successful.

3.2 | Pilot process political success: What was the political impact of the initiation of the program?

The political impact of the initiation of the Vanguard program must be seen in the context of the contemporary political environment. When the Five Year Forward View was published in 2014, the NHS, along with other public services in the United Kingdom and elsewhere, was subject to so-called "austerity" policies (The Centre for Local Economic Strategies, 2014) designed to support recovery from the global financial crash of 2008. NHS funding was consequently growing slowly and below the level of health care cost inflation (Appleby & Gainsbury, 2017). The NHS was predicting a significant funding shortfall (Torjesen, 2012), and NHS England and the Department for Health and Social Care were negotiating with HM Treasury for additional funding. This was obtained in the form of "sustainability and transformation" funding, a proportion of which was used to support the Vanguard program (NHS England, 2014b). This funding came with an expectation that there would be associated "transformation" in service delivery, rather than being used to pay down deficits. The Vanguard program was thus established at least in part to "frame the future" (Nair & Howlett, 2016, p. 1) by demonstrating that the NHS could change, and it could do so rapidly. Our respondents picked this up:

> In practice, what we found, a very strong interest, very often politically driven, to start demonstrating results very quickly. And so, suddenly there'd be reports, the Secretary of State wants an update every Monday morning on rates of non-elective admissions in Vanguard areas, versus other areas. Well, hang on a minute. That's not how the program's supposed to be up and running, and within a year you're starting to ask those questions. (ID018)

While funding was initially offered to all Vanguard sites, by the third year of operation ongoing funding was tied to performance against centrally-determined targets, with Vanguards required to show that they had "earned their way" (NHS England, 2017, p. 47). In November 2017, just over a year into the program, a speech made by the Chief Executive of NHSE England (NHSE) (https://fabnhsstuff.net/fab-stuff/simon-stevens-ceo-nhs-england-speech-nhs-providers-birmingham-november-8th-full) argued that the "Five Year Forward View 'recipe' is working," before asserting that the main problem facing the NHS is not excess demand but "fragmentation and funding." The speech

finished with a suggestion that to support the further rollout of the beneficial service changes demonstrated by the Vanguards, further additional funding would be required. Thus, the appeal to the Treasury for additional funding over and above that already provided was explicitly linked to the demonstration that the NHS had made rapid beneficial changes in service delivery. The rapid initiation and delivery of the Vanguard pilots was thus used *politically* to argue for additional funding for the NHS, and this argument was successful, with a new 5-year funding deal announced in July 2018 (https://www.gov.uk/government/speeches/pm-speech-on-the-nhs-18-june-2018). Thus, in political terms as we have defined them, relating to the distribution of power or resources between actors in a political system, the pilot process was also successful, with NHS England achieving its objective of obtaining additional funding.

3.3 | Pilot outcome program success: Did the pilot program achieve its stated goals?

In terms of Ettelt et al.'s (2014) characterization of the purposes of pilot programs, policy documents (NHS England, 2014a, 2014b, 2015a, 2015b, 2017) suggest that the Vanguard program was conceived of as supporting both early implementation and learning, with an explicit intention that Vanguard sites would test out approaches to change, which could be spread more widely (Checkland et al., 2019). There was also an element of demonstration, with the argument made that: "All three of these care models [PACS, MCPs and ECH] will demonstrate the reinvention of out of hospital care, with PACS and MCPs organising this for the whole population, and enhanced care homes targeting their approach to a care home setting" (NHS England, 2015b, p. 4).

In July 2015 further guidance set out an explicit goal for the Vanguard program to develop approaches which could be subsequently rolled out more widely:

> Each Vanguard system is rooted in its local diverse community. The national New Care Models programme draws together these individual local threads into explicit patterns, in order to exploit common opportunities for radical care redesign and remove barriers to change. Through the support package, our focus is on creating simple standard approaches and products, based on best practice and co-produced with Vanguards, which are designed from the outset for national spread. (NHS England, 2015b, p. 9)

In addition, the program was established with a well-resourced evaluation program, suggesting that these pilots also embodied an experimental approach with genuine uncertainty as to whether the new approaches would work:

> The new models need to show how they help solve the particular issues confronting that particular health community, with proper safeguards against unintended consequences …. There'll need to be independent evaluation, and regular performance benchmarking against comparable area, with periodic opportunities to decide whether to continue with or amend, the arrangements. (Stevens, 2014)

There were thus a number of program goals:

- To implement integrated care programs in designated Vanguard areas.
- To use Vanguard experiences to design "standard approaches and products" which could be rolled out.
- To monitor performance against "benchmarks" and use this information to guide future investment decisions.

The service changes introduced by Vanguards were eclectic, building upon previous initiatives and existing collaborative relationships. A great deal of activity happened in Vanguard sites, and new services or ways of working were introduced. In pilot outcome terms, there was thus demonstrable program success in initiating and running new services.

However, there was less success in developing the promised "standard approaches and products" to be rolled out widely. The program ran for 3 years, and toward the end of the program "frameworks" for each of the Vanguard types were published (NHS England, 2016b, 2016c, 2016d). Two of these (MCPs and PACS) are at a high level of abstraction and diffuse, indicating areas of work which might be considered by those seeking to better-integrate care across sectors. Neither offers standard approaches or "products" which could straightforwardly support local action. The Enhanced Healthcare in Care Homes framework is somewhat more specific, setting out services which should be provided to improve care in Care Homes (NHS England, 2014a).

Finally, it was intended that performance would be monitored and managed. At the start of the program Vanguards were given significant leeway to determine their own outcome objectives, but halfway through the program this changed, with funding for the final year contingent upon success against two metrics—reducing emergency hospital admissions and reducing the average length of stay in hospital.

The impact of the program on these standardized metrics has been explored in a quantitative evaluation (Morciano et al., 2020). The findings are nuanced, but essentially show that the program was associated with a small fall in emergency admissions toward the end of the program, and this was concentrated in the Care Home Vanguards. There was no effect on hospital length of stay.

In summary, in program terms the success of the pilot program was mixed. Following rapid initiation, pilot sites worked quickly to make changes to the way services were delivered. However, the intention to use the program to develop "standard approaches and products" to support rapid replication of the Vanguard "new models of care" was not realized, apart from to a limited extent in the Care Home sites. Finally, by midway through the program, funding was contingent on reducing emergency admissions, a standardized metric of success. There is evidence that this outcome was achieved in a limited way toward the end of the program, mainly in the Care Home sites.

3.4 | Pilot outcome political success: Did the outcomes of the program have political impact?

We will now consider the political impact of the program outcomes, in terms of our definition of "political" as relating to the distribution of power and resources within the UK state and from the perspective of those initiating the pilots. In 2019 NHS England published its NHS Long Term Plan (NHS England, 2019b). This 10 year plan set out NHS priorities over the longer term and was, in part, a response to the announcement that spending on the NHS would increase at more than the rate of inflation for the next 5 years. The Plan sets out how the NHS will spend this "taxpayers' investment" responsibly (NHS England, 2019b, p. 100). While not explicitly asking for further additional funding, the plan could be argued to be part of the ongoing case being made by one public service that it was deserving of an additional share of public resources.

In making this case, the "success" of the Vanguard program forms a prominent element of the argument:

> Following three years of testing alternative models in the Five Year Forward View through integrated care 'Vanguards' and Integrated Care Systems, we now know enough to commit to a series of community service redesigns everywhere. The Vanguards received less than one tenth of one percent of NHS funding, but made a positive impact on emergency admissions, and demonstrated the benefits of proactively identifying, assessing and supporting patients at higher risk to help them stay independent for longer. (NHS England, 2019b, p. 13)

This was accompanied by a bar chart, comparing the rate of growth of emergency admissions between two types of Vanguard (Multispeciality Community Providers and Primary and Acute Care Systems) and 'the rest of the NHS'. The chart is labelled as showing that Vanguards had succeeded in reducing the rate of growth, and it is stated that this finding is statistically significant.

The source of the chart is not stated in the document, but it is likely to derive from an internal evaluation of the Vanguard program which has not been published; it is therefore not available to be examined. Furthermore, the graphic does not provide any details as to how the figures were derived. Notwithstanding this, the existence of a single outcome metric showing a positive effect is used in national policy documents as evidence of pilot program success. The outcomes of the pilot program (as presented in this chart) are thus being used *politically* because they are underpinning the claim that the NHS is deserving of additional resources.

3.5 | Pilot effects program success: What happened next?

Although the Long Term Plan references "a series of community service redesigns everywhere," in practice, only the Enhanced Healthcare in Care Homes Vanguard has been implemented more widely. Linked to a new primary care contract, groups of GPs are being incentivized to set up new services for Care Home residents which have some similarities to the Vanguard ECHC service framework (Coleman, Croke, & Checkland, 2020b). In other areas elements of the Vanguard MCP and PACS service changes have been retained or locally spread, but there has been no systematic wider implementation as envisaged by the initial pilot policy, with no "simple standard approaches" which can straightforwardly spread (Checkland et al., 2021). Thus, there was no clear program success in the form of widespread implementation of new service delivery models derived from the pilots.

One reason for the failure of the pilot program to catalyze the widespread changes originally envisaged is that national policy changed only a year into what was intended to be a 5-year program. This new policy mandated the formation of 44 geographical areas of the country into groups of care providers who were obliged to produce plans (sustainability and transformation plans—STPs) to promote service integration (Hammond et al., 2017). At this point, the pilot projects had yet to be fully implemented, and there were no obvious beneficial outcomes. Yet the STP national policy stated that funding would only be granted to each area if they could address the following questions:

> What are your plans to adopt new models of out-of-hospital care, e.g. Multi-specialty Community Providers (MCPs) or Primary and Acute Care Systems (PACS)? Why should NHS England prioritise your area for transformation funding? And when are you planning to adopt forthcoming best practice from the enhanced health in care homes Vanguards? (NHS England, 2015a, p. 15)

Thus, before any beneficial outcomes were possible, initiating wider rollout was a condition of obtaining additional funding for a different program. However, as we have seen, such rollout has not occurred, apart from in the limited case of Care Homes. Participants in the Vanguards told us that they felt that in the second half of the program policy attention had shifted elsewhere.

> By year 3, attention had moved onto the next shiny thing … (ID013)
> … we expected there to be some kind of conference where all the Vanguard teams would come together, and everybody would say what they'd done, you know, specific to their own team, and you know, what their own statistics were. There was nothing, nothing like that at all. (S4R011)

The program effect success of the pilots has also been attenuated by the failure to realize the initial aim of learning from the pilots' experience. While there was a broad and well-financed internal evaluation program (NHS England, 2016a), and each Vanguard procured a local evaluation (Wilson et al., 2019), no overall report has been published, and there is no public statement of "lessons learned" available. It could thus be said that, despite a rhetorical commitment to drawing together wider lessons to facilitate pilot impact, there were only limited attempts to do this.

TABLE 6 Summary assessment of the "success" of vanguards as pilots

	Program success	Political success
Pilot process	Full success—rapidly and successfully initiated	Rapid initiation of change program used politically to bolster arguments for additional funding
Pilot outcomes	Pilots successfully implemented and locally popular Some outcome goals met Standardized "models of care" only developed in the care of Care Home Vanguards	Outcome success against a single metric used politically to support a longer-term policy program
Pilot effects	Rollout limited to Care Home Vanguard, although elements from other models used locally to inform integrated care developments Limited evidence of systematic learning from the pilots	Claims made that new initiatives are based upon "learning from the Vanguards"; pilots used to make arguments about new legislative change

In summary, there has only been limited success in terms of longer-term pilot program effects, with only a small degree of rollout to non-pilot areas (in respect of Care Homes) despite announcements and policy requirements that this should happen.

3.6 | Pilot effect political success: What has been the political effect of the pilot?

It is perhaps too early to clearly identify the extent to which the overall effect of the pilots supports political claims to success. As evidenced by the Long Term Plan and narratives around sustainability and transformation partnerships there is a political narrative that the Vanguard program has "worked" and has shown how services should be redesigned to improve integration as well as demonstrating the efficient use of "taxpayers' investments," but there is limited evidence that this has led to any specific subsequent activity, with the exception of new services for Care Homes. Policy focus is now upon the transformation of sustainability and transformation partnerships into what are known as "Integrated Care Systems." In policy documents setting how these are being developed, the Vanguard program is again referenced as providing a blueprint:

> [Integrated Care Systems] also incorporate learning from initiatives such as the 50 'vanguards' that tested and refined new care models. In the most successful of these vanguards, NHS providers and commissioners, councils, care homes and others developed more preventive approaches to care and saw significant reductions in emergency admissions. (NHS England, 2019a, p. 2)

Thus, it is claimed that Vanguard learning has been incorporated into new policy, in spite of limited publicly-available evidence that this is actually the case. While this new development is not overtly political in the sense of bolstering arguments for an additional share of resources, the Integrated Care Systems policy has a political element in that their establishment will require legislative change (NHS England, 2020). Thus, claims to lasting effects arising from the Vanguard pilots are being used to support calls for particular changes to legislation. Moreover, Vanguards received considerable additional funding (National Audit Office, 2018), whereas follow on initiatives have not. This embeds a political (distributional) inequity in the developing system.

There is thus some evidence of ongoing political dividends and effects associated with the pilot program, with ongoing legislative developments predicated upon claims of pilot success.

4 | DISCUSSION

We began this paper by suggesting that the conceptualization of policy pilots as rational, experimental processes proceeding in discrete stages is not reflected in the reality which can be messy, performative, and political. Recognizing this, drawing conclusions about the success or failure of pilots becomes more complex and uncertain than the assessment of prespecified outcomes from standardized interventions. Building upon others' work in this field, we have developed a framework to support deconstruction of the impact of policy pilots across a number of dimensions and exploration of success within each. Applying this framework to the Vanguard program case study, we have found that it supports a more nuanced, detailed account of different aspects of pilot "success." This facilitates moving beyond a simple assessment of whether or not initial outcomes were met to interrogate the ways in which the pilot program has been used politically to achieve other things. This fits with Ettelt et al.'s (2014) account of the complexity, ambiguity and mobility of the purposes of policy piloting, providing a framework which surfaces pilot impacts which might otherwise be hidden, and potentially allowing more nuanced causal explanations to be considered.

Table 6 summarizes our assessment of the "success" of the Vanguard pilot program against three categories within two dimensions: program and politics.

In our case, while claims have been made that current policy is building upon lessons learned from the Vanguard program, there is little public evidence of any systematic attempt to draw lessons from the pilot program. One possible explanation for this may lie in the strongly positive political claims to success which have been made nationally. Once the Long Term Plan had declared the program a "success," a more nuanced study of what had gone badly as well as what had gone well becomes unnecessary and perhaps more difficult, with the danger that public consideration of problems or difficulties might disturb the narrative of success constructed to bolster the case for additional NHS funding.

This assessment is from the perspective of the body initiating and running the pilots, NHS England. Space precludes a full assessment from other perspectives as recommended by McConnell et al. (2020), but it is possible to see that this exercise could be rerun from alternative viewpoints. For example, exploration from the perspective of local participants in the Vanguard program might lead to consideration of the local and national political advantages for both individuals and organizations arising out of association with a high-profile pilot program. In this vein, Bailey et al. (2017) suggest that local actors' reputations were enhanced by association with local pilots which were seen to have influenced national policy, while Hammond et al. (2021) found that, regardless of the lack of any meaningful local program success in high-profile innovation policy pilots, local actors felt that their engagement with the pilot program positioned them well for further funding opportunities. Alternatively, examination from the perspective of the Department of Health and Social Care might suggest a political dividend arising out of the appearance of supporting a rational approach to healthcare reform, something seen as valuable in the aftermath of what was generally agreed to be a disastrous major reorganization of the NHS in 2012 (Timmins, 2012). The explicit declaration of the perspective from which the framework is being applied facilitates this type of engagement with the multi-scalar and temporally and geographically bounded nature of policy pilots.

The model we have proposed extends the work of McConnell and Marsh (Marsh & McConnell, 2010; McConnell, 2010) and Bovens (2010). In particular we have shown how assessments of pilot success can usefully separate out an assessment of the outcomes of particular pilots from the assessment of the longer-term effect of the pilots in influencing policy more generally, either via wider rollout or via judicious adjustment of policy design, engaging with the temporal aspect of policies as advocated by t'Hart (2019). Moreover, we have shown that such longer-term effects have both program and political dimensions. However, the political dynamics associated with the Vanguard program are particular in the sense that NHSE, an arm's-length body, is driving policy change while simultaneously making the case for the health service to receive additional funding (Hammond et al., 2018; Rutter, 2014). While the process, outcome, and effect elements, both in relation to programmatic success and political success, are features that can clearly be ascribed to any policy pilot, future research could usefully explore the application of the framework to pilots in other sectors and contexts to explore avenues for its refinement and to consider its wider

applicability. We would argue that our broad definition of "political success" and our explicit use of multifaceted evaluation approaches supports potential cross-sector transferability, but this contention should be tested.

Perhaps the most complex area of the framework is in judging political success. We have judged political success to be evidenced by the making of claims in other contexts which suggest that this particular policy pilot program was successful or important in order to bolster arguments or support other policies which aim to improve the funding status of the NHS or influence the legislative agenda. However, we have not engaged with a broader consideration of political effects in terms of the advancement of the interests of other relevant parties, nor of the possibility that pilots may address altogether more complex political goals or hidden agendas (McConnell, 2018). For example, it is possible that, on occasion, simply doing something, regardless of outcome, may act to reduce political pressure—so-called "placebo" policies (McConnell, 2020). We do not see evidence that the Vanguard pilots fall into this category, but this illustrates the fact that judgments of political success must take a broad and expansive view beyond ostensible or clearly visible effects.

Time is central to the character of pilots, which imply some transitory constellation of actors and elements intended to foster the development of some more permanent form (Bailey et al., 2019). Our proposed framework facilitates the separating out of pilot outcomes from wider effects, including learning or rollout, and allowing the analyst to consider local success against stated goals separately from longer-term effects. However, when objectives are malleable and change during the pilot period as with the Vanguard program, then a process for drawing conclusions about success needs to make explicit the answer to "success as defined when?" and clearly chart the revision of objectives and claims of success, and the political or process consequences associated with any of these (t'Hart, 2019). In the case of the Vanguards, we have suggested that the early declaration of success may have had important consequences for wider policy. Moreover, in keeping with Ettelt et al. (2014), shifting objectives also had process implications for the pilots, as an initial permissive approach shifted to a focus on a single metric of success halfway through the program.

5 | CONCLUSION

Policy pilot evaluations often adopt a relatively simplistic approach of considering whether or not a pilot has "worked" against particular outcome criteria. We have shown that this represents a limited understanding of the purposes and effects of pilots. We have brought together literature on policy success with that on policy pilots to generate a framework within which empirical evaluation findings can be synthesized with analysis of the wider policy landscape to consider pilot success in a more nuanced and multifaceted way. While no such framework can be completely comprehensive, and judgments as to "success" in each category will be contingent, provisional, and potentially arguable, we would suggest that the most valuable aspect of the framework is its focus on explicit delineation of different dimensions of success and on the declaration of the perspective being adopted. While different commentators may disagree with particular judgments in each cell of our summary table, the criteria by which we are judging and the specific aspect of the pilot being judged are clear, providing a more nuanced evaluation framework and facilitating constructive discussion. Furthermore, we believe that explicitly considering pilots across all of these dimensions holds promise in supporting the design of more comprehensive and nuanced evaluation programs which move beyond a simplistic attempt to demonstrate "what works." We have also shown how the explicit separation between program and political elements of success allows interrogation of the antecedents of particular policy decisions, facilitating deeper understanding of contemporary policy trajectories. This in turn supports a more nuanced understanding of later policy developments. For example, we suggest that early political claims to the success of the Vanguard program may have inhibited longer-term consideration of the practical steps required to facilitate program success, with potentially significant implications for NHS managers seeking to implement new approaches to service integration. Moreover, we would suggest that our framework facilitates a more nuanced understanding of pilots, looking beyond stated rationales, and encouraging evaluators and others to explicitly

consider the extent to which particular pilots do indeed represent a rational attempt to test policy ideas, or whether they in fact represent an approach to implementation. "Success" can then be judged against these more complex objectives. We invite others interested in this area of research to consider the applicability of this approach in other fields, with a particular focus upon issues of temporality and methods to identify criteria by which political success might be judged.

ACKNOWLEDGMENTS

We are very grateful to Charles Tallack and Samantha Hinks for their help and support during this project, and to our colleagues, Matt Sutton and Marcello Morciano. We are also very grateful to our participants, who were generous with their time.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Data from this study is not available, due to the requirement to respect participant anonymity.

ORCID

Kath Checkland https://orcid.org/0000-0002-9961-5317

Simon Bailey https://orcid.org/0000-0001-9142-2791

REFERENCES

Ansell, C.K. & Bartenberger, M. (2016) Varieties of experimentalism. *Ecological Economics*, *130*, 64–73. https://doi.org/10.1016/j.ecolecon.2016.05.016

Appleby, J. & Gainsbury, S. (2017) Five reasons why the Chancellor must find extra cash for the NHS at the budget [Comment]. Available at: https://www.nuffieldtrust.org.uk/news-item/five-reasons-why-the-chancellor-must-find-extra-cash-for-the-nhs-at-the-budget [Accessed 19th October 2021].

Bailey, S., Checkland, K., Hodgson, D., Mcbride, A., Elvey, R., Parkin, S. et al. (2017) The policy work of piloting: a case study in the English NHS. *Social Science & Medicine*, *179*, 210–217.

Bailey, S., Hodgson, D. & Checkland, K. (2019) Pilots as projects: policy making in a state of exception. In: Hodgson, D., Fred, M., Bailey, S. & Hall, P. (Eds.) *The projectification of the public sector*. Abingdon on Thames, UK: Routledge, pp. 130–148.

Bovens, M. (2010) A comment on Marsh and McConnell: towards a framework for establishing policy success. *Public Administration*, *88*(2), 584–585.

Brodkin, E.Z. & Kaufman, A. (2000) Policy experiments and poverty politics. *Social Service Review*, *74*(4), 507–532. https://doi.org/10.1086/516423

Burch, M. & Wood, B. (1983) *Public policy in Britain*. Hoboken, New Jersey. USA: John Wiley and Sons Ltd.

Checkland, K., Coleman, A., Billings, J., Macinnes, J., Mikelyte, R. & Laverty, L. https://www.necsu.nhs.uk/wp-content/uploads/2019/07/2019-07-InterimReportNCM-ManchesterUniversity.pdf [Accessed October 20th 2021].

Checkland, K., Coleman, A., Croke, S., Billings, J., Mikelyte, R. & Macinnes, J. https://prucomm.ac.uk/august-2021-national-evaluation-of-the-vanguard-new-care-models-programme-report-of-qualitative-case-studies.html [Accessed 20th October 2021].

Coleman, A., Billings, J., Allen, P., Mikelyte, R., Croke, S., MacInnes, J. et al. (2020a) Ambiguity and conflict in policy implementation: the case of the new care models (Vanguard) programme in England. *Journal of Social Policy*, *50*(2), 285–304.

Coleman, A., Croke, S. & Checkland, K. (2020b) Improving care in care homes: what can primary care networks learn from the Vanguards? *Journal of Integrated Care*, *29*, 85–96. https://doi.org/10.1108/JICA-06-2020-0037

Ettelt, S., Mays, N. & Allen, P. (2014) The multiple purposes of policy piloting and their consequences: three examples from National Health and Social Care Policy in England. *Journal of Social Policy*, *44*(2), 319–338. https://doi.org/10.1017/S0047279414000865

Hammond, J., Lorne, C., Coleman, A., Allen, P., Mays, N., Dam, R. et al. (2017) The spatial politics of place and health policy: exploring sustainability and transformation plans in the English NHS. *Social Science & Medicine*, *190*, 217–226.

Hammond, J., Speed, E., Allen, P., McDermott, I., Coleman, A. & Checkland, K. (2018) Autonomy, accountability, and ambiguity in arm's-length meta-governance: the case of NHS England. *Public Management Review*, *21*, 1–22. https://doi.org/10.1080/14719037.2018.1544660

Hammond, J., Bailey, S., Gore, O.Z., Checkland, K., Darley, S., McDonald, R. et al. (2021) The problem of success and failure in public–private innovation partnerships. *Journal of Social Policy*, 1–21. https://doi.org/10.1017/S0047279421000192

Harrison, S. & Wood, B. (1999) Designing health service organization in the UK, 1968 to 1998: from blueprint to bright idea and 'manipulated emergence'. *Public Administration*, *77*(4), 751–768. https://doi.org/10.1111/1467-9299.00178

Heilmann, S. (2008) From local experiments to national policy: the origins of China's distinctive policy process. *The China Journal*, *59*, 1–30.

Jowell, R. (2003) *Trying it out: the role of 'pilots' in policy-making: report of a review of government pilots*. London, UK: Cabinet Office, Strategy Unit.

Kingdon, J.W. (1995) *Agendas, alternatives and public policies*. New York: Harper Collins.

Lasswell, H.D. (1936 [2018]) *Politics: who gets what, when, how*. Auckland, New Zealand: Pickle Partners Publishing.

Marsh, D. & McConnell, A. (2010) Towards a framework for establishing policy success. *Public Administration*, *88*(2), 564–583. https://doi.org/10.1111/j.1467-9299.2009.01803.x

Martin, S. & Sanderson, I. (1999) Evaluating public policy experiments: measuring outcomes, monitoring processes or managing pilots? *Evaluation*, *5*(3), 245–258.

McConnell, A. (2010) Policy success, policy failure and grey areas in-between. *Journal of Public Policy*, *30*(3), 345–362. https://doi.org/10.1017/s0143814x10000152

McConnell, A. (2018) Hidden agendas: shining a light on the dark side of public policy. *Journal of European Public Policy*, *25*(12), 1739–1758. https://doi.org/10.1080/13501763.2017.1382555

McConnell, A. (2020) The use of placebo policies to escape from policy traps. *Journal of European Public Policy*, *27*(7), 957–976. https://doi.org/10.1080/13501763.2019.1662827

McConnell, A., Grealy, L. & Lea, T. (2020) Policy success for whom? A framework for analysis. *Policy Sciences*, *53*(4), 589–608. https://doi.org/10.1007/s11077-020-09406-y

Morciano, M., Checkland, K., Billings, J., Coleman, A., Stokes, J., Tallack, C. et al. (2020) New integrated care models in England associated with small reduction in hospital admissions in longer-term: a difference-in-differences analysis. *Health Policy*, *124*(8), 826–833.

Nair, S. & Howlett, M. (2016) Meaning and power in the design and development of policy experiments. *Futures*, *76*, 67–74.

National Audit Office. (2018) *Developing new care models through NHS Vanguards*. Available at: https://www.nao.org.uk/report/developing-new-care-models-through-nhs-vanguards/

Newman, J. (2014) Measuring policy success: case studies from Canada and Australia. *Australian Journal of Public Administration*, *73*(2), 192–205. https://doi.org/10.1111/1467-8500.12076

NHS England. (2014a) *Five Year Forward View*. Leeds, UK: NHS England. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [Accessed 20th October 2021].

NHS England. (2014b) *The forward view into action: planning ahead for 2015/16*. Leeds, UK: NHS England.

NHS England. (2015a) *Delivering the forward view: NHS planning guidance 2016/17 to 2020/21*. Leeds, UK: NHS England.

NHS England. (2015b) *The forward view into action. New care models: update and initial support*. Leeds: NHS England.

NHS England. (2016a) *Evaluation strategy for the new care model Vanguards*. Leeds: NHS England.

NHS England. (2016b) *New care models: integrated primary and acute care systems (PACS)—describing the care model and the business model*. Leeds, UK: NHS England.

NHS England. (2016c) *New care models: the framework for enhanced health in care homes*. Leeds, UK: NHS England.

NHS England. (2016d) *New care models: the multispecialty community provider (MCP) emerging care model and contract framework*. Leeds, UK: NHS England.

NHS England. (2017) *Next steps on the Five Year Forward View*. Leeds, UK: NHS England.

NHS England. (2019a) *Designing Integrated Care Systems (ICSs) in England*. Leeds, UK: NHS England.

NHS England. (2019b) *The NHS Long Term Plan*. Leeds, UK: NHS England.

NHS England. (2020) *Integrating care: Next steps to building strong and effective Integrated Care Systems across England*. Available at: https://www.england.nhs.uk/wp-content/uploads/2020/11/261120-item-5-integrating-care-next-steps-for-integrated-care-systems.pdf

Nyumba, T.O., Wilson, K., Derrick, C.J. & Mukherjee, N. (2018) The use of focus group discussion methodology: insights from two decades of application in conservation. *Methods in Ecology and Evolution*, *9*(1), 20–32. https://doi.org/10.1111/2041-210X.12860

Rogers-Dillon, R.H. (2004) *The welfare experiments: politics and policy evaluation*. Stanford, USA: Stanford University Press.

Rutter, J. (2014) The Quango conundrum revisited—why the government still needs a more coherent approach to arm's-length bodies. *The Political Quarterly*, *85*(2), 148–152. https://doi.org/10.1111/1467-923X.12083

Sabatier, P.A. (2006) *Policy change and learning: an advocacy coalition approach (theoretical lenses on public policy)*.Boulder, Co., USA: Westview Press.

Sabatier, P.A. & Weible, C.M. (2014) *Theories of the policy process*. Boulder, Co., USA: Westview Press.

Sanderson, I. (2002) Evaluation, policy learning and evidence-based policy making. *Public Administration*, *80*(1), 1–22. https://doi.org/10.1111/1467-9299.00292

Stevens, S. (2014) *Speech to the NHS confederation* [Press release].

Tassey, G. (2013) Innovation in innovation policy management: the experimental technology incentives program and the policy experiment. *Science and Public Policy*, *41*(4), 419–424. https://doi.org/10.1093/scipol/sct060

t'Hart, P. (2019) *Great policy successes*. Oxford, UK: Oxford University Press.

The Centre for Local Economic Strategies. (2014) *Austerity uncovered: final report prepared by The Centre for Local Economic Strategies*. Presented to TUC. December 2014.

Timmins, N. (2012) Never again? The story of the Health and Social Care Act 2012. In: *A study in coalition government and policy making*. London: The King's Fund.

Torjesen, I. (2012) NHS is unlikely to meet Nicholson challenge to deliver £20bn in efficiency savings, says King's fund. *BMJ: British Medical Journal*, 345 Available at: http://www.bmj.com/content/345/bmj.e6496.abstract [Accessed 20th October 2021].

Wilson, P., Billings, J., Macinnes, J., Mikelyte, R., Welch, E., Richie, R. et al. (2019) *Investigating locally commissioned evaluations of the NHS Vanguard programme*. Available at: https://www.research.manchester.ac.uk/portal/en/publications/investigating-locally-commissioned-evaluations-of-the-nhs-vanguard-programme(3c8cfbf6-52de-4639-b715-2df5627c105e)/export.html#export

Winton, S. (2013) Rhetorical analysis in critical policy research. *International Journal of Qualitative Studies in Education*, *26*(2), 158–177. https://doi.org/10.1080/09518398.2012.666288

How to cite this article: Checkland, K., Hammond, J., Coleman, A., Macinnes, J., Mikelyte, R., Croke, S., Billings, J., Bailey, S., & Allen, P. (2023). "Success" in policy piloting: Process, programs, and politics. *Public Administration*, *101*(2), 463–480. https://doi.org/10.1111/padm.12790