The art and science of non-evaluation evaluation

Jones, Lorelei

Journal of Health Services Research and Policy

DOI:
10.1177/1355819618779614

Published: 01/10/2018

Peer reviewed version

Jones, L. The art and science and non-evaluation evaluation. Journal of Health Services Research and Policy 2018;23(4):262-267. doi: 10.1177/1355819618779614.

The art and science of non-evaluation evaluation

Lorelei Jones, Department of Applied Health Research, UCL

Abstract

This essay considers some limitations of programme theory evaluation in relation to health care policies. This approach, which seeks to surface 'programme theories' or construct 'logic models', is often unable to account for empirical observations of policy implementation in real-world contexts. I argue that this failure stems from insufficient theoretical elaboration of the social, cultural, and political dimensions of healthcare policies. Drawing from institutional theory, critical theory and discourse theory, I set out an alternative agenda for policy research. I illustrate the issues with respect to programme theory evaluation with examples from my experience of research on large-scale strategic change in the English NHS.

Introduction

Forms of programme theory evaluation are diverse[1], yet unified by research designs that begin by eliciting the ideas, the planning or the "programme theory" that has gone into making an intervention and then move to a period of sustained empirical investigation to gauge the extent to which the theories are met in practice'.[2]

One form of programme theory evaluation, realist evaluation, has become particularly popular within health services research. There are websites, webnars, workshops, training material, toolkits, conferences, email lists, and hashtags dedicated to the approach[3,4,5]. And, following concerns that some people may not be doing it right, there are now guidelines that specify what needs to be reported and how, exactly, this should be done.[6] This essay stems from my growing unease at this tendency, within health services research, to 'thingify' methodology, and the way this may hinder our ability to generate insights relevant to policy and practice. It also comes from a belief that the contribution of social scientists to policy is not just that we know about methodology, we know about social phenomena. And while it behoves us to reassure colleagues, and a range of different audiences, that our research is rigorous, and our accounts are credible, knowledge of social phenomena does not come from strict adherence to methodological guidelines, but from collecting data to test, build on, and complement the existing conceptual and empirical literature.

I proceed by discussing some limitations of programme theory evaluation, in practice, in relation to healthcare policy, based on my own experience. I argue that these limitations stem from the way programme theory evaluation neglects the social, cultural and political features of policy. I then set out an alternative agenda for policy research, using illustrations from my own research on large-scale strategic change in the English NHS. My thesis is simple: there is value to research that is more broadly conceived, that uses social science theory to account for 'what is going on' in empirical contexts, and then draws out the implications for policy and practice.

Looking for a logic model in a haystack

I illustrate some of the limitations of programme theory evaluation, in the context of healthcare policies, with my experience of doing an evaluation of a national policy in England (to promote more 'integrated care').[7] The stated aim of the policy I was evaluating was to improve the coordination of services from the perspective of the patient.[8] National policy documents defined integrated care as 'person-centered coordinated care' but did not prescribe how this should be achieved in localities.[8] Rather, the intention was to support and encourage localities to develop innovative models of service integration that had emerged through networks of health and social care organisations working in collaboration.

There were 14 localities in the first wave of the programme (known as integrated care 'pioneers') and I did fieldwork in four of these.[7] My enduring impression from fieldwork was that, as I said to colleagues at the time, 'I don't think this policy actually exists'. There were so many different interpretations of the policy, both between localities and within localities, between individuals from different agencies and professional groups. In some instances, 'being a pioneer' was akin to a 'badge' denoting excellence in integrating health and social care. At other times, it was apparent that when participants talked about the policy, they were referring to a discrete work stream, on workforce say, or information technology. In one locality, 'it' was a particular governance arrangement; specifically, a board that brought together all the service leaders across health and social care. In another, it was nothing discrete or concrete; rather, it was an ethos, and the aim was that this ethos would permeate everything that local organisations did.

The policy also changed over time, during implementation, in the process of becoming concrete, and in response to demands from the centre to demonstrate significant cost savings. Whereas initially there was a diversity strategies across the 14 localities, both in terms of espoused goals and planned activities, overtime these converged on a narrow set of concrete initiatives (e.g a community-based multidisciplinary team and a care navigator), and the focus, of both national and local actors, shifted from improving the coordination of care from the perspective of the patient to decreasing the number of emergency admissions and reducing costs.

I became increasingly frustrated as I struggled to complete the pre-specified evaluation tasks. My inability to 'pin down' the policy meant that even seemingly straightforward tasks, such as collecting and collating information on what localities were doing, were problematic. There was difficulty establishing 'what was in and what was out', in terms of activities, as it proved difficult to establish a starting point for the policy, and local actors would often include pre-existing initiatives. There was also a difference between what local actors were doing, and what they said they were doing for the purpose of the evaluation.

A key evaluation task was to develop a logic model for each locality. Logic models provide a graphical depiction of the programme theory that underpins a policy or programme. Following Hills[9], the logic models we produced divided and labelled the local implementation of the policy as follows: context, inputs, outputs, outcomes, and impact. Yet, to my mind, identifying 'a logic model' seemed nonsensical, when what people were doing, and why, was so dynamic; when logic models were multiple and conflicting and incoherent and contested; when they were mobilised as part of the micro-politics of planning; and when they were mobilised for the purposes of evaluation.

The elusive nature of the policy meant that we spent a lot of time as a team producing these logic models. Large amounts of time were spent discussing what belonged where ('is it an output or an outcome?' 'You put that as an outcome? I put that as an input'). Collecting and analysing qualitative data is time-consuming at the best of times, but contracts for policy evaluation, especially formative evaluation designed to feed into the policy process, characteristically involve extremely tight deadlines. A focus on producing logic models risks not just limiting the relevance of the research (in failing to capture the empirical reality of implementation in local settings) but of restricting research effort to description, rather than analysis. I became frustrated because I felt that if I could just do some research that was more broadly conceived I could say something insightful and helpful to policy and practice. I could build on existing knowledge, make links with findings from previous research, and draw on social science theory to provide an account of 'what was going on', and the implications for policy and practice.

Socio-cultural perspectives on policy

A key principle of realist evaluation, in the context of policy research, is that the effects of a policy are mediated by context.[10] One effect of the popularity of this approach has been a profusion of journal papers on the topic of 'context' and numerous frameworks explicating its social dimensions.[11] But it is not just 'context' that is social, so too is policy. I attribute this neglect of the social, cultural and political dimensions of policy to the status of programme theory evaluation as what Flyvberg calls 'science as usual'.[12] Policy is treated as a given, assumed to fixed, stable, and goal-orientated (implicit in the concepts of 'programme theory' and 'logic model'). However, as illustrated above, when it comes to health care policy, none of these may be safe assumptions.

Socio-cultural perspectives on policy reveal its non-rational features. March and Olsen[13], for example, argue that if you look at real-world decision-making through a rational lens it doesn't make sense, it appears 'chaotic'. This is because in real life people make decisions according to roles, routines, rules, duties, interests, ideas and values. In real life, change is often driven less by problems than by solutions, what Cohen et al call 'a solution looking for a problem'.[14] This is captured in a joke John Glasby tells about integrated care: if you called a policy-maker in the middle of the night to tell them your house was on fire, they would say 'have you tried integrated care?'.[15]

Moreover changes, *and the intentions behind changes*, are transformed by the process of change. In other words, plans change over time, but so too does the rationale. The effect, as March and Olsen observe, is that 'it is difficult to describe a decision, problem solution or innovation with precision, to say when it was adopted and to treat the process as having an ending' (p. 63).[13] This feature of policy is well illustrated by the case of large-scale strategic change in the NHS. When I first started studying the NHS in 1997, there were proposals to introduce a 'super hospital', serving a population of up to 2 million people.[16] Since then, local organisations have appeared to be in a state of perpetual planning as mergers, hospital closures and acute care 'reconfigurations' are proposed but rarely implemented before either being abandoned or rolled into the 'next thing'. [17]

Policy is not fixed, or stable, but fluid, 'a set of shifting, diverse, and contradictory responses to a spectrum of political interests', as the political scientist Murray Elderman described it.[18] Policy, as Graham Allison showed in his analysis of the Cuban Missile Crisis[19], is the

outcome of numerous instances of negotiation and bargaining, both major forms, such as between government departments, but also minor forms, between lower-level players, such as over the wording of a memo. The outcomes of these multiple instances of negotiation and bargaining combine with what Allison calls 'foul-ups' (misunderstandings, reticence on the part of players etc.) to form a 'collage' that constitutes government action on an issue.

Policy is also characteristically ambiguous, often intentionally so.[20] It is the ambiguous nature of policy that enables programme leads to align stakeholders. Ambiguity allows stakeholders with very different views of the problem, and what needs to be done, to sign up to a course of action, something researchers should consider when they next call for greater clarity on the underlying logic model. And policy is invariably oversold. The goals of a policy are intentionally bold and ambitious, 'to excite people', encourage organisations to participate, and motivate front-line staff.[21] This, again, questions the value of programme theories or logic models as the basis of evaluation. At the very least, it will come as no surprise to practitioners if a key finding from a policy evaluation is that the policy did not meet the stated objectives.

Some policies are symbolic,[22] their primary purpose to reaffirm certain shared values. In such cases, substantial effort may go in to formulating the policy, with less interest in subsequent implementation. For example, it is plausible that the centre is more concerned with reducing emergency admissions, or otherwise reducing spending, than with the realisation of integrated care as 'person-centred co-ordinated care', especially in the present climate of financial constraint. Conversely, some policies are too sensitive to be articulated.[22] It has been the policy of successive governments in England to take capacity out of the acute sector, [23] yet political controversy over hospital closures means that it is difficult to find this articulated in national policy documents. As Carole Weiss[24] observed in the context of policy development in the USA:

> Given the consequent grandiosity and diffuseness of program goals, there tends to be little agreement, even within the program, on which goals are real – in the sense that effort is actually going into attaining them – and which are window-dressing. With this ambiguity, actors at different levels in the system perceive and interpret goals in different ways. What the Congress writes into legislation as program objectives is not necessarily what the Secretary's office or the director of the national program see as their mission, nor what the state or local project managers or the operating staff actually try to accomplish.

As a result, the researcher may end up 'evaluating the program against meaningless criteria' or accepting 'bloated promises and political rhetoric as authentic program goals'.[24]

Non-evaluation evaluation

Research does not need to be framed as 'evaluation' for it to be relevant to evaluation. I suggest that there is a value to doing research that is more broadly conceived. Programme theory evaluation is a 'top-down' approach to policy research. It establishes the objectives of a policy and then considers to what extent these objectives were accomplished. In contrast, 'bottom-up' approaches begin with the decisions and strategies of local actors, and consider how these shape local policy implementation. A bottom-up perspective 'does not assume that a policy is the only – or even the major – influence on the behaviour of people involved in the

process'. [25] One of the advantages of this perspective is that by beginning with the perceived problems of local actors, and the strategies developed to deal with them, it affords a consideration of the relative influence of different policies and initiatives on the actions of local implementers. The wider lens adopted by bottom-up approaches can capture how different policies interact, and both intended, and *unintended,* consequences. Bottom-up approaches can also provide information related to the unstated goals of policy, such as the political benefits (e.g. how the policy is received by different stakeholders), which may be of considerable interest to decision-makers at all levels.[24]

I also argue for the explanatory and predictive value of social science theory. Programme theory evaluation does not preclude engagement with social science theory, but what I am suggesting is social science theory applied to the more general question of 'what is going on?'. I suggest that of particular relevance to policy research is social science theory that is interpretive (attuned to multiple meanings), critical (attuned to the operation of power), and discursive (attuned to the use of rhetoric, the role of knowledge and expertise, and the effects of discourse). This is because health care policy is implemented in pluralistic contexts where stakeholders can have very different perspectives on the problem and therefore very different ideas about what should be done.[26] However, it is not simply that stakeholders have different perspectives on policy issues; they have different access to resources and occupy different positions in institutional hierarchies. It is therefore not enough for research to be attuned to multiple perspectives; it must also be attuned to power.[27] Implementing policy in pluralistic contexts is an essentially political activity and it is important to understand the strategies adopted by actors for asserting identity, maintaining autonomy, extending jurisdiction, and promoting particular visions of health care. And it is important to understand how this is accomplished through rhetoric.

Studying large-scale strategic change

In this section, I illustrate the value of research that is more broadly conceived with my research on strategic change to hospital services.[28] In England, enduring political debates about how to best organise hospital services stem from the interests of different stakeholders and the multiple, and at times conflicting, objectives of tax-funded public services. They also reflect the different meanings that hospitals have for different social groups. So, for example, while the overriding concerns for managers may be the efficiency and operational effectiveness of the service, from the perspective of community groups, hospitals are about more than just health services: they are places of considerable social and symbolic significance.[29] For much of the last 50 years, regional-level planners in England have sought to rationalise hospital services. The 'solution' (centralising hospital services into fewer, larger units) has, over time, been attached to a range of different 'problems', including the need for cost savings, the need to provide more responsive services, the need for greater clinical effectiveness, and, more recently, 'safety'.

My research on large-scale change emerged while I was doing fieldwork for another study which involved attending local planning meetings. During these meetings, I noticed that local health service managers would refer to the 'clinical case for change' when presenting plans to close hospital services. The argument was that closing the hospital or hospital department was based on the evidence and necessary for safety. I noticed this, because it was something I kept hearing, and because it contradicted earlier planning documents that contained a financial rationale for change. I was also struck by the rhetorical force of this rationale. It

appeared to me to have the effect of 'shutting down' debate, which I thought was significant given the emphasis in national policy on involving the public in decisions about service delivery. The extent and implications of this observation then became the focus my study.

I found that, at both national and local levels, plans, and the rationales behind plans, were constantly shifting. Drawing on theories of rhetoric and discourse,[30,31] I interpreted the shift in the underlying programme theory, from a financial to a clinical argument, as an instance of strategic reframing, in the context of community resistance to closing hospitals and a concurrent national policy of patient and public involvement in healthcare planning. A key part of this strategy was the enrollment of senior medical leaders to 'sell' the plans to the public.

As rhetoric, the strategy was unsuccessful in that other stakeholders, such as local clinicians and members of the public, were unconvinced by the argument for change. Moreover, these stakeholders perceived the framing, and the enrollment of clinical leaders, as strategic and manipulative. I identified a more insidious form of power in the way that defining the issue as a clinical problem disguised its political nature. Framing the issue as 'clinical necessity' restricted consideration of alternative courses of action, and undermined public participation. These discursive effects were very 'real'. For example, in one instance, local commissioners were able to avoid public consultation by arguing that the proposals were clinical 'best practice'.

The findings from my study problematize many current orthodoxies of policy implementation, such as 'choose the right framing' and 'engage medical leaders'.[32] It shows how, in practice, these strategies are co-optive devices, and suggests that, to the extent that such strategies are recognised by other stakeholders and perceived as manipulative, they may 'backfire' by eroding trust. As Hajer has shown, it is trust, rather than 'evidence', that is key to the acceptance of change in policy controversies.[33] Importantly my study illustrates how programme theory evaluation can fail to capture how programme theories constitute political framings, that shift over time, that are mobilized in political contests, and that have (very real) discursive effects. This lack of fit between the assumptions underlying programme theory evaluation, and what actually happens in real-world situations, limits its relevance and usefulness to policy and practice.

Conclusions

My intention is not to deny the contribution of programme theory evaluation, rather this essay is a call to the research community, particularly those of us who contribute to funding specifications, for greater recognition of the value of research that is more broadly conceived. In the context of complex health care policies, a more general orientation to what is being accomplished, whether organisation is moving is a fruitful direction, and the conditions that support this, can be highly productive in terms of policy learning (see for example Chalabi et al. [34]). Research that is more broadly conceived can suggest different questions and new perspectives, which may be enough to 'dissolve' what were previously seen as intractable implementation problems.[35] At times, a different approach may also be more suited to the needs of decision-makers, at all levels, by providing rapid insight[36], or stakeholder analysis.[37] The leader of the National Institute of Health Research, and Chief Scientific Advisor to the Department of Health, has recently reaffirmed the value to decision-makers of social analysis.[38] A sophisticated analysis of politics, in particular, is of practical benefit, not just in showing what actually happens in the real world, but because of the way politics can

support the implementation of complex social policies that involve multiple, conflicting, yet interdependent, values.[39]

References

1. Blamey A and Mackenzie M. Theories of change and realistic evaluation: peas in a pod or apples and oranges? *Evaluation*. 2007; 13: 439-55.
2. Manzano A and Pawson R. Evaluating deceased organ donation: a programme theory approach. *Journal of Health Organization and Management*. 2014; 28: 366-85.
3. Centre for Advancement in Realist Evaluation and Synthesis.
4. Methods Lab. Realist Impact Evaluation.
5. International Conference for Realist Research, Evaluation and Synthesis: From promise to practice. Brisbane Convention and Exhibition Centre2017.
6. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J and Greenhalgh T. RAMESES II reporting standards for realist evaluations. *BMC medicine*. 2016; 14: 96.
7. Erens B, Wistow G, Mounier-Jack S, et al. Early Evaluation of the Integrated Care and Support Pioneers. London: Policy Innovation Research Unit, 2015.
8. National Collaboration for Integrated Care and Support. Integrated care and support: Our shared commitment. 2013.
9. Hills D. Logic Mapping: Hints and Tips for Better Transport Evaluations. London, 2010.
10. Pawson R and Tilley N. *Realistic evaluation*. London: Sage, 1997.
11. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA and Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*. 2009; 4: 50.
12. Flyvberg B. *Making Social Science Matter*. Cambridge: Cambridge University Press, 2001.
13. March J and Olsen J. *Institutions rediscovered*. New York: Freedom Press, 1989.
14. Cohen M, March J and Olsen J. A Garbage can model of organizational choice. *Administrative Science Quarterly*. 1972; 17: 1-25.
15. Glasby J. If integration is the answer, what was the question? What next for English health and social care. *International journal of integrated care*. 2016; 16.
16. Authorities NAoH. Reinventing Health Care. Towards a New Model. . London: National Assoication of Health Authorities, 1993.
17. Jones L and Exworthy M. Framing in policy processes: A case study from hospital planning in the National Health Service in England. *Social Science & Medicine*. 2015; 124: 196-204.
18. Edelman M. *Constructing the political spectacle*. Chicago: Chicago University Press, 1988.
19. Allison G. *Essence of a decision*. Boston: Little Brown and Company, 1971.
20. Baier VE, March JG and Saetren H. Implementation and ambiguity. *Scandinavian Journal of Management Studies*. 1986; 2: 197-212.
21. Broer T, Bal R and Pickersgill M. Problematisations of complexity: On the notion and production of diverse complexities in healthcare interventions and evaluations. *Science as culture*. 2017; 26: 135-60.

22. Yanow D. Toward a policy culture approach to implementation. *Policy Studies Review*. 1987; 7: 103-15.
23. Gaynor M, Laudicella M and Propper C. Can governments do it better? Merger mania and hospital outcomes in the English NHS. *Journal of Health Economics*. 2012; 31: 528-43.
24. Weiss CH. Where politics and evaluation research meet. *Evaluation Practice*. 1993; 14: 93-106.
25. Elmore R. Backward Mapping: Implementation Research and Policy Decisions. *Political Science Quarterly*. 1980; 94: 601-16.
26. Denis J-L, Lamothe L and Langley A. The dynamics of collective leadership and strategic change in pluralistic organizations. *Academy of Management journal*. 2001; 44: 809-37.
27. Fay B. *Social theory and political practice*. London: Allen and Unwin, 1975.
28. Jones L. Policy as discursive practice: an ethnographic study of hospital planning in England. London School of Hygiene & Tropical Medicine, 2016.
29. Jones L. What does a hospital mean? *Journal of health services research & policy*. 2015; 20: 254-6.
30. Schön D and Rein M. *Frame reflection. Toward the resolution of intractable policy controversies.* New York: Basic Books, 1994.
31. Suddabury R and Greenwood R. Rhetorical strategies of legitimacy. *Administrative Science Quarterly*. 2005; 50: 35-67.
32. Best A, Greenhalgh T, Lewis S, Saul JE, Carroll S and Bitz J. Large-system transformation in health care: a realist review. *The Milbank Quarterly*. 2012; 90: 421-56.
33. Hajer M. *The Politics of Environmental Discouse*. Oxford: Oxford University Press, 1995.
34. Chalabi Z, Erens B, Hajat S, et al. Evaluation of the Implementation and Health Related Impacts of the Cold Weather Plan for England 2012. London Policy Innovation Research Unit, 2015.
35. Zuiderent-Jerak T, Strating, M., Nieboer, A., Bal, R. Sociological refigurations of patient safety; ontologies of improvement and 'acting with' quality collaboratives in healthcare. *Social Science and Medicine*. 2009; 69: 1713-21.
36. Johnson GA and Vindrola-Padros C. Rapid qualitative research methods during complex health emergencies: A systematic review of the literature. *Social Science & Medicine*. 2017; 189: 63-75.
37. Langley A and Denis J-L. Beyond evidence: the micropolitics of improvement. *BMJ Quality and safety*. 2011; 20(Suppl1): i43-i6.
38. Whitty CJ. What makes an academic paper useful for health policy? : BioMed Central, 2015.
39. Nevile A, Kay A and Carey G. Value choices in a mixed economy of care: How politics shapes the implementation of complex social policies. *Social Policy & Administration*.