

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



Ettelt, S; Mays, N (2015) RCTs: How compatible are they with policy-making? *British Journal of Healthcare Management*, 21 (8). pp. 379-382.

Downloaded from: <http://researchonline.lshtm.ac.uk/2305326/>

DOI:

#### Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact [researchonline@lshtm.ac.uk](mailto:researchonline@lshtm.ac.uk).

Available under license: Copyright the author(s)

### **RCTs - how compatible are they with contemporary health policy-making?**

**Stefanie Ettelt and Nicholas Mays**

#### **Abstract**

Randomised controlled trials (RCTs) have been promoted as a means of improving policy-making by testing new policies. While testing before full-scale roll-out is commendable, this paper discusses the challenges of using RCTs in contemporary (national) health policy-making in England. There are at least two challenges in particular that are currently underrepresented in the debate: The first arises from the complexity of many policies which are often too diffuse and unclear in focus to allow for the clear distinction between a policy ‘mechanism’ and its context to be drawn that is required for a RCT. The second challenge relates to the timing of RCTs, which tend to take place either too early in the life of a policy to be meaningful or too late to have an effect on policy formulation. We therefore encourage policy-makers and researchers to be clear about the types of uncertainties ‘field experiments’ are meant to address which may be addressed better by other types of knowledge generation.

**Published in: British Journal of Healthcare Management 2015, 21 (8): 379-382.**

#### **Introduction**

There is a prominent rhetoric permeating English public policy discourse advocating for more frequent use of ‘field experiments’. Recent uses of the term suggest that by ‘experiment’ proponents usually mean randomised controlled trials (RCTs) though ‘experiments’ tend to signify different things to different interests (Oakley 2006). RCTs are an established research design to produce robust evidence on cause and effect relationships (Campbell and Stanley 1963, Rychetnik, Frommer et al. 2002). They are well established in clinical effectiveness research and have over time been employed in a wide array of research disciplines in search of robust measures of outcomes from interventions, programmes and policies, including in health services, public health and education research, international development, and political science (McKee, Britton et al. 1999, Shadish, Cook et al. 2002, Stoker 2010, Picciotto 2012).

In recent years, RCTs have found their way into United Kingdom government rhetoric with reports published by the Cabinet Office promoting RCTs to policy-makers as a means of implementing evidence-based policy-making (Haynes, Goldacre et al. 2012). More recently, Simon Stevens, the NHS England chief executive, has advocated “local experiments” to test and adapt models of care (Stevens 2014). While testing before full-scale roll-out is a valuable proposition in its own right, this paper highlights the challenges of using RCTs and related research designs in health service policy-making in this way.

## **Why RCTs appeal to policy-makers**

While RCTs are still relatively rare in English health policy evaluation (as opposed to their use in the evaluation of drugs and procedures), the argument in favour of RCTs is remarkably persuasive. Its particular appeal to policy-makers is that generally RCTs are regarded as the most robust method of establishing relative effectiveness (strong on “internal validity”). They also convey a sense of certainty and generate confidence in the evaluation and, by extension, it seems, the policy. RCTs also shield policy-makers from the criticism of academics that they are insufficiently ‘evidence-based’ in their decision making, and in some areas, such as international development, they play into the desire to ensure that Government spending is accounted for rigorously (Patton 2008). Robust data of effectiveness are a key ingredient of cost-benefit analyses and cost effectiveness studies, thus in high demand by Government economists (Picciotto 2012). In health policy, RCTs can be particularly persuasive to the medical community and can therefore be a handy strategy to increase the credibility of a new policy within the National Health Service (NHS).

## **Methodological and implementation challenges of conducting RCTs in health policy**

The challenges of carrying out RCTs have already been discussed extensively (MRC 2008) and will only be mentioned here briefly since the main challenges of using RCT evidence relate to the policy process. There are both methodological and practical challenges in conducting policy RCTs. Researchers of various stripes have argued that while RCTs use randomisation to reduce bias when comparing outcomes between intervention and control groups, this approach cannot do justice to the complexity of many public policies (e.g. Pawson and Tilley 1997, Hawe, Shiell et al. 2004, Marchal, Westhorp et al. 2013). RCTs either rely on the researchers’ ability to identify variables *ex ante* to control for complexity or they have to be embedded in a larger programme of work in which other components, such as qualitative studies, make up for their shortcomings. Either way, the argument goes, the complexity of public policy requires ever larger studies. RCTs’ strength in terms of internal validity limits the transferability and generalizability of their findings which, in turn, raises questions about their suitability for public policy evaluation if used in isolation (Cartwright 2007, Cartwright 2013).

Other challenges arise in the implementation of RCTs. Some RCT have suffered from local resistance, specifically when local implementers take issue with using randomisation. Although much design work has gone into devising variations to RCTs that help to assuage concerns about randomising individual patients including cluster randomisation or stepped wedge designs (MRC 2008), the recent experience of the Whole System Demonstrators (WSD) RCT of telehealth and telecare has shown that the expectations of local implementers of the aims of such experiments can differ widely from those of researchers and policy-makers. Local service managers are more likely to argue the case for research that sheds light on the best strategies for bringing about long-term service improvement rather than research that tests different approaches in head-to-head experimental comparisons with a risk

of discontinuation of entire programmes if they are not found to be superior to controls in a trial (Hendy, Chrysanthaki et al. 2012).

Despite these difficulties, scholarly debate on the merits of RCTs seems to have arrived at a point where it is agreed that the methodological and practical challenges of conducting RCTs in health policy can be addressed through careful design and implementation strategies, including embedding the RCT in a wider programme of work and making explicit efforts to improve the communication of the rationale of RCTs to patients, managers, the media and the public (Oakley, Gough et al. 2005, Oakley 2006, Bonell, Fletcher et al. 2012). However, there are other obstacles that are much less discussed and not so easily resolved.

### **RCTs and the policy process**

The process of policy-making in English central Government poses obstacles to an effective use of RCTs. Evidence-informed policy is not only about commissioning and conducting evaluations, and summarising their findings for policy advisers, but also about the ability to use such studies in decision-making. Proponents of RCTs argue in a straightforward way that more robust evidence of effectiveness from RCTs is more likely to have an impact on policy makers and their decisions (Haynes, Goldacre et al. 2012), yet there is little empirical evidence to support this assumption (Cook 2007).

There are at least two main challenges to be overcome in using RCTs in making policy decisions. The first is that RCTs rely on a definition of policy as an ‘intervention’ while conceptions of policy are often much more diffuse and, deliberately or unwittingly, unclear in focus. There is a well-established argument about complexity in evaluation. For example, Hawe and colleagues have argued convincingly that many policies in public health resemble “a time limited series of events, new activity settings and technologies that have the potential to transfer the system because of their interaction with the context and the capability created from this interaction” rather than discrete interventions that are “replicable and recognisable ... in each site” (Hawe, Shiell et al. 2004, Hawe, Shiell et al. 2009). In a similar vein, Pawson maintains that “programmes are complex systems thrust into complex systems” (Pawson 2014). This lack of easily identifiable mechanisms undermines the evaluability of such policies through RCTs as it is not possible to isolate cause and effects from contextual factors that cannot be controlled for. Multi-centre RCTs may increase the transferability of the findings, but again this is limited to the number of sites involved and may not help in understanding the complex interplay of mechanism and context.

In other cases, it may not be possible to evaluate the effectiveness of a policy at the early stage when policy makers may most need guidance from evaluation because the policy has yet to mature to the extent that it can be replicated - its ‘mechanism’ is still emergent and local study sites will inevitably vary in how they interpret the content of the policy. This is particularly relevant if pilots (perhaps more appropriately termed ‘trailblazers’) are used to promote policy development locally; the implementation of pilots (i.e. the so called experiment) then becomes part of the policy development process rather than part of policy testing. The issue here then is timing, assuming that the content of policy becomes clearer as

the experiment progresses, although participants may not be aware of this lack of conceptual clarity at the beginning of the experiment.

However, local experimentation can also be a conscious strategy to promote policy development, as the Direct Payment in Residential Care Trailblazer programme illustrates (which does not include an RCT). In such cases, local variation is seen as part of the strategy for policy learning through a process of local trial and error. Again, it is unlikely that such an approach to experimentation could produce the kind of standardised intervention that RCTs require to come to meaningful conclusions. Indeed, RCTs can undermine the strategic approach to encouraging local policy innovation, as the RCT design by definition requires a high degree of researcher control and standardisation that may be counterproductive to local development and learning (Hendy, Chrysanthaki et al. 2012).

The second challenge is that it is unclear when in the policy process RCTs of policies should be conducted (Ogilvie, Cummins et al. 2011). If RCTs as policy experiments are used at the policy formulation stage, it will in many cases be too early in the life of the policy if the ‘mechanism’ of policy is yet to be developed, as argued above. If the purpose of the RCT is to inform policy implementation, it may be too late to make significant changes to the policy if findings suggest that the policy is not (sufficiently) effective or cost effective. Policy-makers, be they politicians or officials, are likely to have already committed themselves to a direction of travel and will be loath to admit that their judgement was made in error. They will not be rewarded for doing so in the adversarial style of policy-making in England. The feedback loop of public, media and political actors promotes a policy-making style that puts particular value on decisive action creating another incentive for policy-makers to commit to policy at an early stage and to continue to adhere to a policy despite the evidence.

These reflections raise a dilemma for policy evaluation. Assuming it is possible to identify ‘an intervention’ within the policy, it may still be too early in the life of the policy/intervention to conduct an RCT at policy formulation stage, as the policy content may not be sufficiently ‘mature’. However, by the time the policy is mature enough to be tested using an RCT it may well be too late in the life of the policy for the findings to inform implementation, as assumptions of effectiveness have already found their way into legislation and other forms of codification of policy, let alone into local practice. Equipoise, a key principle of RCTs in clinical practice, is unlikely to survive the political pressures characteristic of policy-making despite the wishful thinking of some researchers (Petticrew, McKee et al. 2013).

## **Ways forward**

This is not to throw the baby out with the bathwater and advocate that RCTs have no place in health policy evaluation. However, it is worth giving more attention to the challenges posed by the reality of the policy-making process, discussed above, if ‘experiments’, conceptualised as RCTs, are to be used and useful in health policy-making.

By way of recommendation, we encourage policy-makers and researchers involved in this debate to consider a wider notion of experimentation, which allows for a more flexible approach to choosing research designs for evaluation. The craft and art will be for researchers and policy makers to strive better to understand the nature of the policy that is to be tested and the contextual factors likely to shape its effects. Notions of ‘field experiments’ organised as RCTs seem to rest on a misunderstanding of the types of uncertainty surrounding policies, which are often more about how best to operationalize a particular policy and how to ensure its compatibility with existing systems and previous policies than about the effectiveness of the new policy compared with the status quo ante in any straight forward sense. RCTs, however, address the second type of uncertainty far better than the first. A potential way of employing RCTs that reflects the dominant policy uncertainty from the policy makers’ perspective would be to randomise ‘sites’ to one of a range of different approaches to operationalising the policy in question rather than attempting to use the RCT to compare the (new) policy with the status quo.

Ultimately, however, whether RCTs have a place or not in health policy evaluation, it is for policy-makers to make informed judgements about the type of knowledge they require to make better policy decisions, and for evaluators to advise them in this pursuit using a range of types of research rather than for both groups to persist in over-estimating what RCTs can achieve in health policy evaluation.

## References

- Bonell, C., A. Fletcher, M. Morton, T. Lorenc and L. Moore (2012). "Realist randomised controlled trials: A new approach to evaluating complex public health interventions." Social Science & Medicine **75**: 2299-2306.
- Campbell, D. T. and J. Stanley (1963). Experimental and quasi-experimental designs for research. Chicaco, Rand McNally College Publishing.
- Cartwright, N. (2007). "Are RCTs the gold standard?" BioSocieties **2**(1): 11-20.
- Cartwright, N. (2013). "Knowing what we are talking about: why evidence doesn't always travel." Evidence & Policy: A Journal of Research, Debate and Practice **9**(1): 97-112.
- Cook, T. D. (2007). "Describing what is Special about the Role of Experiments in Contemporary Educational Research?: Putting the "Gold Standard" Rhetoric into Perspective." Journal of MultiDisciplinary Evaluation **3**(6): 1-7.
- Hawe, P., A. Shiell and T. Riley (2004). "Complex interventions: how "out of control" can a randomised controlled trial be?" BMJ: British Medical Journal **328**(7455): 1561.
- Hawe, P., A. Shiell and T. Riley (2009). "Theorising interventions as events in systems." American journal of community psychology **43**(3-4): 267-276.
- Haynes, L., B. Goldacre and D. Torgerson (2012). Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. London, Cabinet Office.
- Hendy, J., T. Chrysanthaki, J. Barlow, M. Knapp, A. Rogers, C. Sanders, P. Bower, R. Bowen, R. Fitzpatrick and M. Bardsley (2012). "An organisational analysis of the implementation of telecare and telehealth: the whole systems demonstrator." BMC Health Services Research **12**(1): 403.
- Marchal, B., G. Westhorp, G. Wong, S. Van Belle, T. Greenhalgh, G. Kegels and R. Pawson (2013). "Realist RCTs of complex interventions—an oxymoron." Social Science & Medicine **94**: 124-128.
- McKee, M., A. Britton, N. Black, K. McPherson, C. Sanderson and C. Bain (1999). "Interpreting the evidence: choosing between randomised and non-randomised studies." Bmj **319**(7205): 312-315.
- MRC (2008). Developing and evaluating complex interventions: new guidance. [www.mrc.ac.uk/complexinterventionsguidance](http://www.mrc.ac.uk/complexinterventionsguidance), accessed 29 May 2015. London, Medical Research Council.
- Oakley, A. (2006). "Resistances to 'new' technologies of evaluation: education research in the UK as a case study." Evidence & Policy: A Journal of Research, Debate and Practice **2**(1): 63-87.
- Oakley, A., D. Gough, S. Oliver and J. Thomas (2005). "The politics of evidence and methodology: lessons from the EPPI-Centre." Evidence & Policy: A Journal of Research, Debate and Practice **1**(1): 5-32.

- Ogilvie, D., S. Cummins, M. Petticrew, M. White, A. Jones and K. Wheeler (2011). "Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers." Milbank Quarterly **89**(2): 206-225.
- Patton, M. Q. (2008). Utilization-focused evaluation. Los Angeles, SAGE Publications.
- Pawson, R. (2014). "Are randomised control trials essential in policy making?" SRA Research Matters **2**: 1.
- Pawson, R. and N. Tilley (1997). Realistic evaluation. London, SAGE.
- Petticrew, M., M. McKee, K. Lock, J. Green and G. Phillips (2013). "In search of social equipoise." BMJ **347**(7918): 18-20.
- Picciotto, R. (2012). "Experimentalism and development evaluation: Will the bubble burst?" Evaluation **18**(2): 213-229.
- Rychetnik, L., M. Frommer, P. Hawe and A. Shiell (2002). "Criteria for evaluating evidence on public health interventions." Journal of epidemiology and community health **56**(2): 119-127.
- Shadish, W. R., T. D. Cook and D. T. Campbell (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, Houghton Mifflin Company.
- Stevens, S. (2014). Speech to the NHS Confederation Annual Conference 2014. Liverpool 4 June.
- Stoker, G. (2010). "Exploring the Promise of Experimentation in Political Science: Micro-Foundational Insights and Policy Relevance." Political Studies **58**(2): 300-319.