The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance

Warren Pearce · Sujatha Raman

Abstract Since 2012, a new movement of government departments, think tanks and high-profile individuals within the UK has sought to promote the increased usage of randomised controlled trials (RCTs) in public policy. They promote RCTs as an evidence-based corrective for the inappropriate certainties of experts. Recent government reports and public debate around this initiative are reviewed and analysed within a framework for epistemic governance: normative insights into how knowledge for policymaking should be understood and governed drawn from science and technology studies and the policy sciences. The legitimacy of RCT evidence within policymaking is found to rest on the recognition of three key features: (1) how multiple meanings of evidence limit generalisability, (2) ensuring a plurality of evidence inputs, including those from other forms of research and expertise, and (3) building institutions for governing the use of RCTs in the public interest. Producing evidence for policymaking is a hybrid activity that necessarily spans both science and politics. Presenting RCTs as naively neutral evidence of what policy interventions work is misleading. The paper concludes by calling for more work on how the new RCT movement might engage with its own history in social and policy research on the value of experiments for policymaking.

W. Pearce (✉)
Institute for Science and Society (ISS), University of Nottingham, Nottingham NG7 2RD, UK
e-mail: warren.pearce@nottingham.ac.uk
URL: http://www.nottingham.ac.uk/sociology/people/warren.pearce

S. Raman
Leverhulme Making Science Public Programme, Institute for Science and Society (ISS),
University of Nottingham, Nottingham, UK
e-mail: sujatha.raman@nottingham.ac.uk
URL: http://www.nottingham.ac.uk/makingsciencepublic

S. Raman
Visiting Scholar, Consortium for Science, Policy and Outcomes, Arizona State University, Tempe,
Arizona , USA

Introduction

A renewed movement promoting the use of randomised controlled trials (RCTs) in public policy has emerged in the UK, organisations such as NESTA (Puttick 2011), Education Endowment Foundation (Torgerson and Torgerson 2013), Behavioural Insights Team (Haynes et al. 2012) and the high-profile doctor, academic and author, Ben Goldacre (2011) highlight the beneficial role RCTs have played in challenging entrenched tendencies to favour the expert judgment of clinicians over systematic examination of the evidence. In a paper for the Department for Education, Goldacre argue that the use of RCTs can similarly benefit public policy by offering evidence of 'what works' rather than relying on 'eminence, charisma, and personal experience' (2013a, p. 8). Following this renewed interest, the Department for Education (2013) announced it was running two RCTs to research the impact of a child protection assessment tool on school attainment in mathematics and science. This and other UK government papers advocating increased usage of RCTs across public policy (Haynes et al. 2012), and particularly in the education sector (Goldacre 2013a), has prompted extensive debate among teachers, education researchers and other commentators (Allen 2013; James 2013; Whitty 2013; Goldacre 2013b).

In this paper, we examine the promises and pitfalls of this movement, drawing on insights at the interface of science and technology studies (STS), political theory and the policy sciences on the role of scientific evidence and expertise in policymaking (Brown 2009; Guston 2001; Hisschemöller and Hoppe 1995; Hoppe 2005; Jasanoff 2003a, b, 2012; Miller 2008; Wesselink and Hoppe 2011; Wynne 2010). This extensive literature has been reviewed and synthesised elsewhere (e.g., Hoppe 2005; Hoppe et al. 2013), so here we focus on key normative principles for the problem of 'epistemic governance' (Raman 2015). Epistemic governance suggests that the production of knowledge for governance *itself needs be governed*, as STS and policy scholars have long highlighted. Recent RCT advocates argue that the evidence gathered from these trials is a better, more reliable form of knowledge than the personal knowledge of experts. But such knowledge is insufficient if it is abstracted from the political side of policymaking in which problems are framed and knowledge given meaning (Wesselink and Hoppe 2011). This does not mean that the normative objective of how best to think about and produce knowledge for policy is irrelevant. Rather, it requires a rethinking of this knowledge production as distinctively hybrid, as both scientific and political (Hoppe et al. 2013; Jasanoff 2012; Raman 2015). This paper explores the implications of this insight for the RCT movement, focusing specifically on its recent incarnation in policy practice.[1]

RCTs are advocated as an effective way of determining whether or not a particular intervention has been successful at achieving a specific outcome. A population is randomly divided into two or more groups and given different interventions. The results within each group are then compared at the end of the trial period. The element of randomisation is intended to eliminate the influence of external factors on these results; for example selection bias (Haynes et al. 2012, pp. 8–9). Such trials are commonplace in medicine, but have also been conducted in areas of public policy within the UK, including charity giving,

---

[1]   Some political scientists have also begun to advocate 'experiments' as a way of learning what works for policy (Stoker and John 2009; Stoker 2010), but for reasons of space, we bracket this literature.

voter turnout, recycling (John et al. 2011), domestic energy efficiency (Heyman et al. 2011) and sex education in schools (Oakley et al. 2006).

Concerns about RCTs include the fear that these are being promoted as the 'gold standard' in a hierarchy of evidence that marginalises qualitative research (Cartwright 2007), the neglect of issues of interpretation and meaning in the desire to tame complexity with numbers (Greenhalgh 2012), the de-professionalisation of experts, and the failure to consider ethical questions around random assignment of potentially beneficial interventions. By contrast, RCT advocates (e.g., Goldacre 2013a) argue that the method can empower professionals by setting them free from the edicts of government ministers, and that trials are ethically acceptable in situations where people have little idea of whether an intervention is beneficial, and that critics fail to consider the potential harm and economic cost of interventions not being tested before being widely implemented.

In fact, the debate on RCTs within the UK is not new (Macintyre and Petticrew 2000), although this history is overlooked by the new movement emerging from the Cabinet Office (Haynes et al. 2012). Indeed, the distinguished social researcher Ann Oakley (1998a) traces the origins of the method of random assignment to the sociology of education in the early twentieth century which pre-dates its development and use in medicine. Oakley and others such as Donald Campbell, the pioneer of 'quasi-experimental' research methods in policy evaluation, have also set out a case for why and how RCTs might benefit the public (Oakley 1998b; Campbell and Stanley 1963; Macintyre and Petticrew 2000). However, it is not clear how well the lessons from this previous era, such as the need to incorporate qualitative insights from professional and 'lay' users or those impacted by the research, have been taken on board by the contemporary RCT movement.

Our analysis is presented in three sections. First, we draw on insights from STS and policy sciences to outline normative principles for epistemic governance developed in depth elsewhere (Raman 2015) to assess the credibility of evidence in public policy. Second, we describe the case made by the new RCT movement through recent UK government reports, locating these recent trends within the context of prior work on the role of applied social research such as RCTs and quasi-experiments in policymaking. Third, we re-examine the claims of the new RCT movement through the lens of epistemic governance, outlining its potential promise and pitfalls.

The normative problem of epistemic governance

Arguments for increased RCT usage as a means of suppressing the influence of experts appear reminiscent of STS analyses of how the uncertainties which surround evidentiary claims are often not made explicit (Stirling 2012; Wynne 1992). For example, scientists are frequently called upon to advise on matters of potential risk from existing or new technologies and, in turn, policy questions about how these issues should be governed. In these debates, it is often assumed that scientists will provide the facts on risk which policymakers should then use to make decisions about regulation. Yet, the 'factual' assessments of scientists have been contested in cases such as nuclear power, pesticides, and genetically modified organisms and a number of contingencies that shape their assessments have been made more explicit through the course of controversy. Importantly, some of these challenges emerge from within science itself.

For example, looking at the controversy around the risks of mobile phones in the late 1990s–early 2000s, Stilgoe (2007) highlights the response from a new group headed by Sir William Stewart (former Government Chief Scientific Advisor) that was asked to take a

fresh look at an issue that had been traditionally handled by the National Radiological Protection Board (NRPB). In the course of this, uncertainties that NRPB had taken as unproblematic became problematised by the Stewart group as matters needing further scrutiny (Stilgoe 2007). If there was little scientific work done on long-term exposure to electromagnetic fields or impacts on potentially vulnerable groups, what then was the basis of NRPB's reassuring messages about the safety of mobiles? The new Stewart group recommended a precautionary approach in this context. On a first reading, it appears that fresh, and 'better', evidence helped reveal the kind of 'inappropriate certainty' of experts critiqued by proponents of RCTs (Goldacre 2013a).

However, what this argument ignores is that technical evidence is still marshalled, interpreted and made relevant to policy by experts. There is no getting away from the fact that those advocating RCTs are still experts. Given the fact that there are multiple forms of specialised and less-codified 'lay' knowledge of relevance to policy, the fundamental question then is: who counts as an expert or what counts as expertise? Similarly, we need to ask: what does and does not count as evidence?

In this context, STS research suggests that the normative challenge is not so much the use of expertise but the way it is exercised by technical experts who wish to offer evidence to policymakers (e.g., in the RCT case, evidence of what interventions 'work'). In particular, experts acting in an official capacity as science advisors or in partners with government need to clarify how they:

(a) Draw on a range of knowledge sources, attending to underlying assumptions in that knowledge and its limits.
(b) Handle different epistemic and ethical concerns.
(c) Synthesise multiple inputs, including from other experts and the public, to form their judgments.

When experts seem to fall short in this admittedly demanding task, new evidence is unlikely to resolve the problem, as that evidence will require its own expert interpretation and assessment. The core problem, therefore, is one of epistemic governance: how the production of evidence for policymaking is, and should be, governed. We highlight three principles synthesised from the science and policymaking literature by Raman (2015) that we then use to assess the new RCT movement. These principles complement existing frameworks of boundary arrangements between science and policy (Hoppe 2005) that depend on the nature of the policy problem at stake; in particular, the extent to which problems are more or less well structured in terms of epistemic and value-based claims (Hisschemöller and Hoppe 1996). Where these frameworks extensively map different actual and normatively desirable options in relation to political context, the purpose of Raman's (2015) synthesis of principles is simpler. Its aim is to clarify the qualities sought from expertise when neither scientism (leaving decisions to scientists) nor decisionism (leaving decisions to policy makers) nor populism (leaving decisions to 'the public') is normatively satisfactory, and policy problems are inevitably complex. The three principles are: evidence possesses multiple meanings, there are plural sources and types of evidence, and hybrid institutions are required to manage evidence's inherent complexity.

Multiple meanings

Appeals to evidence in policymaking and debate are frequently met with further questions on what sort of evidence counts, how the evidence is to be interpreted, what evidence is credible and importantly, how the policy question is defined and framed in the first place.

Thus, evidence is always mediated—its meaning(s) not self-evident (Lasswell 1948, p. 218). Evidentiary claims rest on tacit assumptions made in a specific context, so their transferability to another context is in question. Even if evidence is uncontested, its role in addressing policy questions is open to negotiation as the way these questions are framed can be re-opened. For example, in the case of risk-based regulation of new technologies such as genetically modified organisms, Wynne (2010) has argued that evidence of risk is often assumed to be the central issue at stake (to be settled by science) when public concerns may relate to other matters such as the future of intensive agriculture and innovation or the choices made in research funding priorities. Thus, appeals to evidence-based policymaking are misleading if they strip out the context for interpreting specific forms of evidence.

Plurality of evidence

Evidence-based policy (EBP) does require expert advice to be opened up, but this goes beyond Prime Minister Cameron's stated desire of 'bringing information out into the open' as a means to 'hold government…to account' (Prime Minister's Office 2010). Rather it means being open to, and engaging with, a plurality of forms of evidence and value judgments (Stilgoe et al. 2006). This includes acknowledging differences within science as well as between science and public meanings. Sarewitz (2004, p. 388) highlights an 'excess of objectivity' within science, meaning that scientific disciplines may parse 'nature' differently with different criteria as to what counts as evidence—for example, molecular biology and ecological sciences approach the potential hazards of genetically modified organisms differently. Again, the case for evidence-based policy often ignores the multiplicity of evidence relevant to the policy problem in question.

The role of institutions

The turn to evidence may be a welcome sign in public policymaking (Miller 2008; Jasanoff and Winickoff 2002), but it requires institutions to be considered trustworthy in their handling of the plurality of evidence, the meaning and framing of evidence and the limits of scientific knowledge at any point in time. No single expert can be expected to handle this complexity—hence the need for institutions. Different political systems have developed different 'civic epistemologies' (Miller 2008) and 'boundary arrangements' (Hoppe 2005) in terms of how they decide the nature and role of knowledge and criteria for evidence, make sense of epistemic conflict and the mechanisms for resolving it. The expert advisory committee is one such institution for distilling evidence for policy purposes, but problems arise when it is seen as—and claims to be—doing 'science' or providing unmediated evidence, as opposed to a hybrid activity that is simultaneously political and scientific (Guston 2001; Wesselink and Hoppe 2011).

We now consider how problems of expertise are represented in the new RCT movement, before assessing the extent to which RCT advocates acknowledge the significance of context and meaning in evidence, the need to engage with plural forms of knowledge and the need for institutions to mediate these challenges.

The new RCT movement as a response to problems of expertise and political ideology

Ben Goldacre is a key figure in the drive to increase the use of RCTs in public policy, co-authoring a report on the potential for the method's usage with members of the UK

government's Behavioural Insights Team (Haynes et al. 2012) as well as discussing and popularising it through blogs (Goldacre 2012a), national radio (Tzabar 2013) and web tools (Puttick 2013). Goldacre draws on evidence-based medicine (EBM) as a guide to developing EBP, claiming that EBM replaced an over-reliance on expertise with a focus on data:

> when trials were first introduced in medicine, they were strongly resisted by some clinicians, many of whom believed that their personal expert judgement was sufficient to decide whether a particular treatment was effective (Haynes et al. 2012, p. 13)

> policy people need to have a little humility, and accept that they don't necessarily know if their great new idea really will achieve its stated objectives. (Goldacre 2012a)

> *inappropriate certainty* can be a barrier to progress, especially when there are charismatic people, who claim they know what's best, even without good evidence. Medicine suffered hugely with this problem, and as late as the 1970s there were infamous confrontations between people who thought it was important to run fair tests, and "experts", who were angry at the thought of their expertise being challenged, and their favourite practices being tested. (Goldacre 2013a, p. 12, emphasis added)

Here, expertise is depicted as part of the problem, typified by a cabal of professionals who are set in their ways and opaque to outside scrutiny. Over-reliance on such expertise is shown to have been harmful within certain cases in medicine; for example, the persistent practice of using steroids to treat brain injuries was shown to harm, rather than help, patients only through evidence derived from a series of RCTs (Haynes et al. 2012, pp. 16–17; CRASH trial collaborators 2005). So, RCTs are seen as a way of dealing with the edge of expert judgement, testing whether or not an intervention works in the way that it is claimed to. This is seen as a lesson from EBM which can be transferred into public policy.

In its own terms, the case for RCTs seems to address powerful critiques of the inaccessibility of expert judgment to those outside specialist circles. Its rhetoric is reminiscent of what Fuller (1997) described as one of the earliest notions of science—a form of democratic knowledge accessible to all irrespective of background or specialised training. Ezrahi (1990, p. 286) highlights that accountability in liberal democracies has traditionally been based on a framework of what he calls an 'attestive visual culture in which actions are treated as observable factual events' with discernible causes that can be perceived or monitored in the public sphere. Though Ezrahi goes on to argue that this is no longer possible in a mediatised world where symbols are more important than attestation of facts, RCTs represent a promise to fulfil older expectations of providing public facts for democratic monitoring of government activity. As well as responding to the limitations of expertise, RCTs appear to be an effective response to criticisms that of the UK government's policymaking is purely ideological and unfettered by evidence (e.g., Priestland 2013). Thus, at first sight, RCTs seem to offer a corrective to decisionist models of rule in which politicians dominate as well as technocratic models in which experts dominate (Hoppe 2005; Weingart 1999).

These arguments were brought into sharp focus with the publication of Goldacre's paper (2013a) for the Department for Education on increasing the evidence base in education, which was launched with comments by Secretary of State Michael Gove (Department for Education 2013). This development led to a mixed reception from education researchers, teachers and other onlookers (Allen 2013; Goldacre 2013b (comments); James 2013;

Whitty 2013). Critics noted that Gove's apparent to commitment to EBP was not supported by a recent controversy over changes to the history curriculum made with little reference to expert opinion (Cannadine 2013). Goldacre used the antipathy of many teachers to Gove as an argument for RCTs, arguing that increasing their usage would make teaching 'a truly independent profession', with teachers enabled to carry out their own research and provide the evidence necessary to affect policy decisions (Goldacre 2013b).

In this vision, evidence appears as a way of challenging policy decisions that are otherwise based simply on a problematic political ideology. The normative aspiration seems clear, yet it contradicts a previous statement by Goldacre where the relationship between politics and evidence appears far more tame:

> You use your ideology to set your policy objectives and the moral and ethical limits of what you are willing to do (Goldacre 2012b)

While this appears a sensible riposte to those fearing that RCTs represent technocratic government trumping political and ethical frameworks for decision-making (Archer 2011), it also disavows the normative objective of marshalling evidence to improve policymaking by re-drawing a clear line between expertise and politics in a decisionist fashion. It concedes that if evidence challenges the core values underpinning a political administration, it will constitute the kind of 'uncomfortable knowledge' which will be hard to assimilate into policy (Rayner 2012) and that nothing much can be done about this. The notion that expertise for policymaking is a hybrid activity requiring engagement with politics (and political ideology) as well as with science and evidence is conspicuously absent.

The way in which RCTs have been rolled out by the current UK government reinforces this separation and primacy of political decisions from evidence. So while RCTs promise to answer the question of 'what works' from a menu of policy interventions, determining which interventions are to be tested in the first place has been a political decision. Potential policy options regarded as politically unpalatable are unlikely to be tested, while a policy option may be so in tune with prevailing political values that subjecting it to tests is regarded by politicians as unnecessary or even unwelcome. For example, in 2013, the Justice Secretary, Chris Grayling, commented on his decision to introduce a criminal rehabilitation policy of payment by results before the completion of a pilot study:

> The last Government were obsessed with pilots. Sometimes those in government just have to believe in something and do it, but the last Government set out a pilot timetable under which it would have taken about eight years to get from the beginning of the process to the point of evaluation and then beyond. Sometimes we just have to believe something is right and do it, and I assure Members that if they went to Peterborough [the location of the pilot study] to see what is being done there, they would think it was the right thing to do. (Hansard 2013)

Although the pilot study in question was not a RCT, Grayling's comments highlight two areas of tension between the demands of political judgement and those of establishing an evidence base. First, they reassert the primacy of political judgement over evidence; if a democratically elected representative believes something is 'the right thing to do', then he or she has the power and formal authority to act, irrespective of the evidence base. One may also locate Grayling's decision to roll out payment by results as influenced by the managerial tradition of governance which has become prevalent over the last two decades (Rhodes 2011, pp. 27–28).[2] Second, it shows the potential for temporal disjoint between

---

[2] Grayling spent a number of years working as a management consultant (Conservatives n.d.).

establishing evidence and demands for political action. Faced with what he perceived to be an urgent problem, Grayling highlighted the length of time that a pilot study takes to be undertaken and evaluated, which runs contrary to the need to exercise political authority in a mediatised environment (Hajer 2010). Trials and pilot studies which stretch beyond the life of a parliament are potentially exposed to a change in political leadership and political values which may find the policy options being tested unpalatable, and opt instead to test interventions more in tune with their world view.

This discussion suggests that the relationship between evidence and political judgment remains tricky. In theory, the legitimacy of decisions taken by elected politicians may be opened up if new forms of evidence emerge, such as RCTs, which challenge the basis of their judgments. But for this to happen, we need to pay attention to ways in which such evidence might be considered credible. To aid this task, we turn to one of the new RCT movement's forerunners, the Campbell Collaboration.

The Campbell Collaboration emerged from the influential Cochrane Collaboration (itself part of the EBM tradition) which produces systematic reviews of evidence relating to healthcare. Systematic reviews grew out of the observation that amidst the proliferation of scientific studies based on 'single' experiments, there was an urgent need for assessing the body of evidence rigorously in order to assess what it collectively revealed about the question at hand. Reviews are expected to deal with the problem of weeding out poorly designed or inadequately reported studies as well as synthesising results (including conflicting findings) from the rest. This would then provide the best possible answer to questions such as: what is the impact of a specific drug or intervention on patient outcomes? Members of the Cochrane Collaboration felt the need for a similar collective to examine and synthesise evidence on the effectiveness of social interventions. They came together with public policy experts in 1999 to form the international Campbell Collaboration, named in honour of distinguished methodologist Donald T. Campbell, to answer the questions 'What helps? What harms? Based on what evidence?' (Campbell Collaboration, n.d.). The Collaboration's archives are cited within the new RCT movement's key report as a "good starting point…which support policymakers and practitioners by summarising existing evidence on social policy interventions" (Haynes et al. 2012, p. 20)

The Campbell Collaboration argues that Campbell:

> "[A]dvocated the idea that governmental reforms can be seen as societal experiments to which scientific rules of evidence can be applied. He believed that scientific evidence could be generated to estimate the effects of governmental reforms, resulting in better informed policy and practice and ultimately improving people's well being". (Campbell Collaboration, n.d.)

Such a description presents Campbell as adopting a naively positivist perspective on the relationship between evidence and policy. However, Campbell adopted a more plural approach, arguing that 'quasi-experiments' were a more appropriate form of research for public policy purposes, acknowledging that the kind of 'science' carried out in this context was necessarily shaped by real world rather than laboratory conditions (Dunn 1998). While Campbell believed that evidence could and should improve policymaking, a closer reading of Campbell's extensive body of work[3] suggests a sensitivity to the complex social and cultural conditions of evidence-gathering, collaboration and synthesis for policymakers, including 'situation-specific' and 'street' wisdom (Campbell 1984; Dunn 1998). He also made a case for institutions organising the production of social science to encourage interaction between

---

[3] Including as evolutionary epistemologist, philosopher of science and interlocutor of sociology of science.

diverse perspectives and explicitly encourage the production of 'dissenting-opinion research reports' as well as stakeholder participation. Campbell's arguments suggest that, in order to assess the potential of systematic evidence for social and public policy, we first need to put the production of scientific evidence in its social context. These views of a key predecessor to the contemporary RCT movement are consistent with the insights from the STS literature explored in Sect. 2 on the importance of meaning, plurality and institutions in the production of evidence. In the next section, we examine how well the new RCT movement fares in relation to these criteria for enhancing the role of evidence in public policy.

Reassessing the new RCT movement

Multiple meanings

When RCTs are presented as offering generalisable evidence of what works, the conditions and assumptions built into their doing and interpretation are erased from the story. Generalisability is elevated above the context-dependent sensibilities of practitioners who recognise that making institutional apparatus work favourably for clients is not merely a matter of faithfully implementing policy, but making case-by-case decisions which may not fall within recognised policy boundaries (Maynard-Moody and Musheno 2006, pp. 328–329). The wider context in which an intervention works is ignored, and it is implied that success in one context can simply be transferred to another.

The already vexed issue of EBP's relationship to practitioners' ability to make judgements (Greenhalgh and Russell 2009, p. 310) is further accentuated by granting an elevated status to RCTs. There is a clear tension between RCTs claiming to offer unmediated and neutral evidence, and street-level bureaucracy, where practitioners choose to operate beyond official policies in contingent ways (Lipsky 1980). While the act of randomisation can accommodate some of the uncertainties which arise from such practitioner discretion within a particular location, it may be that the character and frequency of such discretionary acts vary greatly between different locations, dependent on hard-to-quantify factors such as organisational culture. So RCTs can control for street-level bureaucracy within a particular location, but it is inappropriate to interpret results as generalisable beyond the limitations of the test itself. Rolling out new policies based on RCTs is then likely to lead to greater levels of deviation from official policy by practitioners on the ground, who remain unconvinced by top-down interventions based on trial results from different locations.

A decisive shift from qualitative to quantitative methods may have longer-term effects on practitioner behaviour and the interpretation of evidence relevant to the public interest. Training programmes move away from a logic of professional, client-focused practice towards a managerial adherence to "rule-governed consequences" which, in seeking to tidy up uncertainty, instead erodes the meaning of public policy (Greenhalgh 2012, p. 96). Jasanoff (1991) has shown elsewhere that where systems for governing knowledge rely on highly formal, rule-based, quantitative methods for resolving uncertainty, they have attracted criticism for 'false precision', failing to make explicit the value judgments that underlie statistical assessment of hazard, and ironically, the failure of strict rules to be sufficiently responsive to new knowledge. In the case she describes, the uncertainties refer to discrepancies between studies employing different methods of estimating carcinogenic risk (e.g., animal-based studies, epidemiological studies), but the point requires broader consideration.

There is also the question of how the policy problem to which RCTs are presented as solutions are defined in the first place. If most policy problems are 'wicked' and perhaps

require 'clumsy solutions' (Rayner 2012), RCT advocates need to engage with different policy framings rather than simply adopt the framing that is already 'given' by the ideology of the time. By presenting RCTs as separate from politics, RCT advocates appear unwilling to engage in this work.

Plurality of evidence

One reading of the claims made by new RCT advocates for public policy is that they are calling for a greater plurality of research methods admitted into policymaking. RCTs, it is said, are underrepresented within public policy, particularly when placed in the context of their contribution to the field of medicine in recent decades. From this perspective, the hostile reaction to Goldacre's education policy paper by some researchers in the field is seen as a defence of methodological turf; an unwillingness to open up their field of research to alternative approaches to research. But the use of RCTs only pluralises in relation to other quantitative approaches that already hold considerable sway in government. Increased use of RCTs reinforces this position, limiting the role of expertise to a technical exercise in reducing uncertainty rather than a practical quest to determine the most appropriate policies using a broad set of criteria (Sanderson 2004, p. 376).

While philosophical debates continue over the status of RCT knowledge claims (Hammersley 2005; Chalmers 2005; Cartwright 2007; Bonell et al. 2012; Marchal et al. 2013), the rhetoric of those proposing greater use of RCTs often goes beyond a plea to merely add RCTs to the methodological toolkit of social researchers. The Cabinet Office report advocating RCTs adopts a clear stance from the outset:

> RCTs are the *best way* of determining whether a policy is working. They have been used for over 60 years to compare the effectiveness of new medicines…. We believe that… [this] approach has the potential to be used in *almost all* aspects of public policy. (Haynes et al. 2012, pp. 6–7; emphasis added)

So while there may be challenges in introducing RCTs in particular cases, the utopian ideal depicted by the Cabinet Office is one not of a multi-perspectival approach to research methodology, but a world where RCTs can be designed and implemented across almost all areas of policy, and are held up as the 'gold standard' for evaluating success and failure (Cartwright 2007). In doing so, the Cabinet Office report appears to downplay guidance from within EBM that RCTs may be unsuitable or implausible for some complex interventions (Craig et al. 2008, p. 26).

Faced with such a vision, it is unsurprising that experienced social researchers feel better disposed towards the status quo than a revolution which relegates non-RCT research to a secondary position, and potentially lead governments to cut funding for those policy areas not amenable to RCTs and are thus seen as non-evidence based. Such scenarios may be a long way removed from current realities (Triantafillou 2013, p. 11). However, by focusing narrowly on RCTs in EBM, advocates have introduced a logic that further cements the dominance of quantitative studies in public policy, relegates qualitative research to a minority role, and impoverishes the policy process.

The role of institutions

As Jasanoff and Winickoff (2002) make clear in their article on the role of evidence in the 2003 Iraq War, for EBP to be meaningful the process by which scientific facts are produced must be transparent: "when transparency and standards break down, science suffers".

Within EBP, there have been calls for government to demonstrate greater transparency in the "evidence audit trail" (Mulgan and Puttick 2013, p. 8). In conjunction with the Coalition Government's open data agenda (HM Government 2012, pp. 14–16; Cabinet Office 2012), RCTs bring the promise of a rich data set which can be analysed and interpreted in multiple ways by multiple actors. The creation of "information architectures" for public policy will enable the kind of step-by-step progress in EBP that is claimed for EBM, learning and adapting from a vast database of RCTs (Goldacre 2013a, p. 16). The template can be found in the Cochrane Collaboration's systematic reviews of trial data. However, as Goldacre himself has highlighted in Bad Pharma (2012c), progress in EBM has been hampered by large pharmaceutical companies failing to disclose trial results which may have negatively impacted on their business interests. Such practices can have significant impacts on public policy. For example, the failure of pharmaceutical company Roche to publish the full results of trials of its Tamiflu drug presented the drug as more effective against influenza than was actually the case, causing the UK government to stockpile a drug which had little discernible effect on sufferers of the illness (Goldacre 2012c, pp. 81–99; National Audit Office 2013). At the time of writing, Goldacre is part of a campaign to shame pharmaceutical companies into publishing the results of all trials in the future, irrespective of the results (Sense About Science 2013).

So proponents of EBM recognise that corporate interests have eclipsed the public interest in the case of publishing RCT results. In Jasanoff and Winickoff's terms, science has suffered from a break down in transparency and standards. The Bad Pharma experience is also pertinent to assessing the wider context of EBP; the Coalition Government's so-called mutualisation of parts of the Civil Service, including the Cabinet Office's Behavioural Insights Team which has driven the RCT agenda for public policy (BBC 2013). Critics dispute the extent to which the new organisations, described as 'social purpose vehicles' (Hood 2013), will operate mutually, pointing out that government mutualisation requires only 25 % of shares to go to staff and that the rights of service users are not protected (Mayo 2013). Such a separation of policy-making function from the state raises the question of how the Behavioural Insights Team can prioritise the public interest, or 'social purpose', in its activities once exposed to market pressures, and how public accountability can be ensured (Dunt 2014). Government mutualisation may risk EBP following the EBM template in a manner its advocates would not wish to see: a metamorphosis from Bad Pharma to 'bad policy' where non-governmental policy makers are driven by commercial interests rather than a public duty to publish the available evidence.

For RCTs to play a role in social science evidence for policy, the data they generate will need to be mediated by institutions at the science/policy interface, for example, an advisory committee that can handle different forms of evidence and give due weight to competing policy definitions in the public interest.[4] Where these bodies are 'mutuals', their legitimacy will rest not just on how they deal with internal organisational matters (ownership) but, crucially, on the way the public interest is addressed. Part of this assessment is the manner in which the RCT movement responds to questions over the ethical dilemmas in assigning potential 'benefits' to one group while another remains in the control group, an issue only skirted over in the Cabinet Office report (Haynes et al. 2012, pp. 16–17).[5]

---

[4] The former Royal Commission on Environmental Pollution (RCEP) was perhaps one such body, though its focus on such questions as giving weight to different forms of evidence led to clashes with more specialist advisory committees centred on expertise from a single scientific discipline (Diggle 2007).

[5] The way this issue is handled is also relevant for the institutional credibility of RCTs.

Conclusion

This paper has provided an overview of the new movement to increase RCT usage in UK public policy. We have analysed this trend in the context of an emerging concern with epistemic governance and shown the benefits of incorporating insights from the STS literature. We have highlighted the tricky relationship between evidence and policy and argued, following Campbell, a key predecessor of the modern RCT movement, that for evidence to be credible it must take account of its specific context—a point that is less apparent in the rhetoric of contemporary promoters of RCTs. In order to clarify the criteria for the legitimacy of evidence in policymaking, we drew on normative principles for epistemic governance (Raman 2015) developed on the basis of literature on the role of scientific evidence in public policymaking. Three insights from this literature have been highlighted as important for the credibility of evidence for public policy: that evidence requires interpretation of meaning and never 'speaks for itself'; that experts providing advice need to acknowledge these meanings and consider a plurality of sources and forms of evidence; and that the institutions have a key role in maintaining transparency and standards in both the production of evidence and its mediation by expert advisors. In short, what Miller (2008) describes as a 'culture of reasoning' within institutions mediating scientific evidence for policy purposes will be an important part of assessing how evidence-based policymaking works in practice.

Asking how well the new RCT movement reflects these criteria for epistemic governance, we have argued that trials hold the promise of increasing the plurality of evidence, but that the rhetoric surrounding the new movement provides a potential pitfall in promoting RCTs at the expense of other research methods. Avoiding this pitfall requires members of the movement to make their case as part of a broader culture of reasoning that is more sensitive to the multiple challenges and dilemmas of policy evaluation in the real world (e.g., the extensive oeuvre of Ann Oakley). Indeed, the EBM history drawn upon by the new RCT movement provides a caution against giving undue weight to any one research method. In particular, the hegemonic status of RCTs within EBM has narrowed, rather than broadened the availability of evidence, with insufficient consideration given to the possibility that complexity cannot be tamed without loss of meaning (Greenhalgh 2012, p. 96). So while RCTs may seek to make expertise more transparent, an over-reliance on them as a research method may make expertise less public in other ways, particularly if they result in an approach to policy and its implementation which increasingly privileges rule-based practices and leaves the public facing an increasingly inflexible and bureaucratic state (Greenhalgh et al. 2014). Finally, RCTs should not be regarded as a counterweight to expertise but as a research method requiring its own set of expert skills to implement and, crucially, to interpret.

There is a danger that the current UK government's interest in RCTs is driven not by their methodological suitability, but because they lend themselves to a model of governance that values context-free quantification and benchmarking. In this situation, RCT advocates would do better by helping build institutions that could put the evidence from trials in its proper context, clarify the conditions under which interventions work or do not work and why, and interpret the meaning of RCTs in relation to plural sources of evidence. This requires engagement across science *and* politics, alongside an acknowledgement that evidence for policymaking requires expertise as well as data. The new RCT movement needs to grasp this message if it is to benefit the lives of those who are the subject of policy interventions.

References

Allen, R. (2013, March 14). *Evidence-based practice: Why number-crunching tells only part of the story*. Retrieved June 18, 2013, from http://ioelondonblog.wordpress.com/2013/03/14/evidence-based-practice-why-number-crunching-tells-only-part-of-the-story/.

Archer, G. (2011, September 2). *Abortion reform: Listen to your prejudices*. *Telegraph.co.uk*. Retrieved May 10, 2013, from http://www.telegraph.co.uk/health/8737694/Abortion-reform-listen-to-your-prejudices.html.

BBC (2013). *"Nudge unit" to be part privatised*. *BBC*. Retrieved from http://www.bbc.co.uk/news/uk-22362727.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science and Medicine, 75*(12), 2299–2306. doi:10.1016/j.socscimed.2012.08.032.

Brown, M. B. (2009). *Science in democracy: Expertise, institutions, and representation*. Cambridge, MA: MIT Press.

Cabinet Office. (2012). *Open Data white paper: Unleashing the potential (Cm 8353)*. Norwich: The Stationery Office.

Campbell Collaboration (n.d.). *Background*. Retrieved June 18, 2013, from http://www.campbellcollaboration.org/background/index.php.

Campbell, D. T. (1984). Can we be scientific in applied social science. Evaluation Studies Review Annual, 9, 26–48.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Cannadine, D. (2013, March 13). The future of History. *Times Literary Supplement*. Retrieved June 18, 2013, from http://www.the-tls.co.uk/tls/public/article1228938.ece.

Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties, 2*(1), 11–20. doi:10.1017/S1745855207005029.

Chalmers, I. (2005). If evidence-informed policy works in practice, does it matter if it doesn't work in theory? *Evidence & Policy: A Journal of Research, Debate and Practice, 1*(2), 227–242. doi:10.1332/1744264053730806.

Conservatives. (n.d.). *The conservative party people members of parliament Chris Grayling MP*. Retrieved May 10, 2013, from http://www.conservatives.com/People/Members_of_Parliament/Grayling_Chris.aspx.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). *Developing and evaluating complex interventions: The new Medical Research Council guidance*. London: Medical Research Council.

CRASH trial collaborators. (2005). Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury—outcomes at 6 months. *The Lancet, 365*(9475), 1957–1959. doi:10.1016/S0140-6736(05)66552-X.

Department for Education. (2013, May 3). *New randomised controlled trials will drive forward evidence-based research*. Retrieved June 18, 2013, from https://www.gov.uk/government/news/new-randomised-controlled-trials-will-drive-forward-evidence-based-research.

Diggle, R. (2007) *Regulatory science and uncertainty in the risk assessment of pesticide residues*, Ph.D. thesis, University of Nottingham http://etheses.nottingham.ac.uk/1451/.

Dunn, W. N. (1998). *The experimenting society: Essays in honor of Donald T. Campbell* (Vol. 11). Transaction Publishers.

Dunt, I. (2014, February 5). Nudge nudge, say no more. Brits' minds will be controlled without us knowing it. *The Guardian*. Retrieved from http://www.theguardian.com/commentisfree/2014/feb/05/nudge-say-no-more-behavioural-insights-team.

Ezrahi, Y. (1990). *The descent of Icarus: Science and the transformation of contemporary democracy*. Cambridge, MA: Harvard University Press.

Fuller, S. (1997). *Science*. Buckingham: Open University Press.

Goldacre, B. (2011, May 14). *How can you tell if a policy is working? Run a trial*. Retrieved September 18, 2014, from http://www.theguardian.com/commentisfree/2011/may/14/bad-science-ben-goldacre-randomised-trials.

Goldacre, B. (2012a). Here's our Cabinet Office paper on randomised trials of government policies. Read it. *Bad Science*. Retrieved from http://www.badscience.net/2012/06/heres-a-cabinet-office-paper-i-co-authored-about-randomised-trials-of-government-policies/.

Goldacre, B. (2012b). *Five Minutes With: Ben Goldacre*. Retrieved May 9, 2013, from http://www.bbc.co.uk/news/science-environment-19517224.

Goldacre, B. (2012c). *Bad pharma: How drug companies mislead doctors and harm patients*. London: Fourth Estate.

Goldacre, B. (2013a). Building evidence into education. *Bad Science*. Retrieved from http://media.education.gov.uk/assets/files/pdf/b/ben%20goldacre%20paper.pdf.

Goldacre, B. (2013b, March 18). Teachers need to drive the research agenda. *The Guardian*. Retrieved from http://www.guardian.co.uk/education/2013/mar/18/teaching-research-michael-gove?CMP=twt_gu&commentpage=1.

Greenhalgh, T. (2012). Why do we always end up here? Evidence-based medicine's conceptual cul-de-sacs and some off-road alternative routes. *Journal of Primary Health Care, 4*(2), 92–97.

Greenhalgh, T., & Russell, J. (2009). Evidence-based policymaking: A critique. *Perspectives in Biology and Medicine, 52*(2), 304–318. doi:10.1353/pbm.0.0085.

Greenhalgh, T., Howick, J., Maskrey, N., & for the Evidence Based Medicine Renaissance Group. (2014, June 13). Evidence based medicine: a movement in crisis? *BMJ, 348*, g3725. doi:10.1136/bmj.g3725.

Guston, D. H. (2001). Boundary organizations in environmental policy and science: An introduction. *Science, Technology and Human Values, 26*(4), 399–408.

Hajer, M. A. (2010). *Authoritative Governance: Policy making in the age of mediatization*. Oxford: Oxford University Press.

Hammersley, M. (2005). Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice. *Evidence & Policy: A Journal of Research, Debate and Practice, 1*(1), 85–100. doi:10.1332/1744264052703203.

Hansard. (2013, January 09): Column 315. *Transforming Rehabilitation*. Retrieved May 10, 2013, from http://www.publications.parliament.uk/pa/cm201213/cmhansrd/cm130109/debtext/130109-0001.htm#13010946001447.

Haynes, L., Service, D., Goldacre, B., & Torgerson, D. (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials*. London: Cabinet Office.

Heyman, B., Harrington, B., Heyman, A., & The National Energy Action Research. (2011). A randomised controlled trial of an energy efficiency intervention for families living in fuel poverty. *Housing Studies, 26*(1), 117–132. doi:10.1080/02673037.2010.512787.

HM Government. (2012). *The Civil Service Reform Plan*.

Hisschemöller, M., & Hoppe, R. (1995). Coping with intractable controversies: The case for problem structuring in policy design and analysis. *Knowledge and Policy, 8*(4), 40–60. doi:10.1007/BF02832229.

Hood, L. (2013, May 29). Nudge unit brings Midas touch to growth agenda. *Research Fortnight*, 4.

Hoppe, R. (2005). Rethinking the puzzles of the science-policy nexus: From knowledge utilization and science technology studies to types of boundary arrangements. *Poiesis & Praxis, 3*(3), 191–215.

Hoppe, R., Wesselink, A., & Cairns, R. (2013). Lost in the problem: The role of boundary organisations in the governance of climate change. *Wiley Interdisciplinary Reviews Climate Change, 4*(4), 283–300.

James, M. (2013, March 20). *New (or not new) directions in evidence-based policy. Response to Ben Goldacre*. British Educational Research Association. Retrieved from http://www.bera.ac.uk/system/files/Mary%20james%20-%20New%20%28or%20not%20new%29%20directions%20in%20evidence-based%20policy.%20Response%20to%20Ben%20Goldacre.pdf.

Jasanoff, S. (1991). Acceptable evidence in a pluralistic society. In D. Mayo & R. Hollander (Eds.), *Acceptable evidence: Science and values in risk management* (pp. 29–47). New York: Oxford University Press.

Jasanoff, S. (2003a). Technologies of humility: Citizen participation in governing science. *Minerva, 41*, 223–244.

Jasanoff, S. (2003b). (No?) Accounting for expertise. *Science and Public Policy, 30*(3), 157–162.

Jasanoff, S. (2012). *Science and public reason*. London: Routledge.

Jasanoff, S., & Winickoff, D. (2002, November 18). Hard facts and soft law: what's the evidence? *openDemocracy*. Retrieved June 13, 2013, from http://www.opendemocracy.net/theme_9-wmd/article_762.jsp.

John, P., Cotterill, S., Richardson, L., Moseley, A., Stoker, G., Wales, C., et al. (2011). *Nudge, nudge, think, think: Using experiments to change civic behaviour*. London: Bloomsbury Academic.

Lasswell, H. D. (1948). *Power and personality*. New York: W.W. Norton & Company.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public service*. New York: Russell Sage Foundation.

Macintyre, S., & Petticrew, M. (2000). Good intentions and received wisdom are not enough. *Journal of Epidemiology and Community Health, 54*(11), 802–803.

Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G., & Pawson, R. (2013). Realist RCTs of complex interventions—an oxymoron. *Social Science & Medicine*, *94*, 124–128. doi:10.1016/j.socscimed.2013.06.025.

Maynard-Moody, S., & Musheno, M. (2006). Stories for research. In D. Yanow & P. Schwartz-Shea (Eds.), *Interpretation and method. Empirical research methods and the interpretive turn* (pp. 316–330). Armonk: M. E. Sharpe.

Mayo, E. (2013, May 3). Nudged out: This is mutualisation, but not as we know it. *The Guardian*. Retrieved from http://www.guardian.co.uk/commentisfree/2013/may/03/nudge-unit-mutualisation-but-not-as-we-know-it.

Miller, C. A. (2008). Civic epistemologies: Constituting knowledge and order in political communities. *Sociology Compass, 2*(6), 1896–1919. doi:10.1111/j.1751-9020.2008.00175.x.

Mulgan, G., & Puttick, R. (2013). *Making evidence useful: The case for new institutions*. London: NESTA.

National Audit Office. (2013). *Access to Clinical Trial Information and the Stockpiling of Tamiflu (Memorandum for the Committee of Public Accounts, HC125 Session 2013–14)*. Norwich: The Stationery Office.

Oakley, A. (1998a). Experimentation and social interventions: A forgotten but important history. *BMJ. British Medical Journal, 317*(7167), 1239.

Oakley, A. (1998b). Public policy experimentation: Lessons from America. *Policy Studies, 19*(2), 93–114.

Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J., & RIPPLE Study Team. (2006). Process evaluation in randomised controlled trials of complex interventions. *BMJ, 332*(7538), 413–416. doi:10.1136/bmj.332.7538.413.

Priestland, D. (2013, April 18). Michael Gove's disdain for experts is typical of the laissez-faire ideologues. *The Guardian*. Retrieved from http://www.guardian.co.uk/commentisfree/2013/apr/18/michael-gove-disdain-experts-typical-ideologues.

Prime Minister's Office. (2010, May 29). PM's podcast on transparency. Retrieved December 3, 2013, from https://www.gov.uk/government/news/pms-podcast-on-transparency.

Puttick, R. (2011). Where is the evidence? *NESTA* Retrieved from http://www.nesta.org.uk/blog/where-evidence.

Puttick, R. (2013, May 28). Randomise Me: How you can set up and run your own trial. *Policy Innovation Blog*. Retrieved from http://www.nesta.org.uk/assets/blog_entries/randomise_me_how_you_can_set_up_and_run_your_own_trial.

Raman, S. (2015). Science, uncertainty and the normative question of epistemic governance in policy-making. In E. Cloatre & M. Pickersgill (Eds.), *Knowledge, Technology and Law: Interrogating the Nexus* (pp. 17–32). London: Routledge.

Rayner, S. (2012). Uncomfortable knowledge: The social construction of ignorance in science and environmental policy discourses. *Economy and Society, 41*(1), 107–125. doi:10.1080/03085147.2011.637335.

Rhodes, R. A. W. (2011). *Everyday life in British Government*. Oxford: Oxford University Press.

Sanderson, I. (2004). Getting evidence into practice: Perspectives on rationality. *Evaluation, 10*(3), 366–379.

Sarewitz, D. (2004). How science makes environmental controversies worse. *Environmental Science & Policy, 7*(5), 385–403. doi:10.1016/j.envsci.2004.06.001.

Sense about Science. (2013, January 9). AllTrials campaign launch. Retrieved June 17, 2013, from http://www.senseaboutscience.org/news.php/292/alltrials-campaign-launch.

Stilgoe, J. (2007). The (co-) production of public uncertainty: UK scientific advice on mobile phone health risks. *Public Understanding of Science, 16*(1), 45–61.

Stilgoe, J., Irwin, A., & Jones, K. (2006). *The Received wisdom: Opening up expert advice*. London: Demos.

Stirling, A. (2012). Opening up the politics of knowledge and power in bioscience. *PLoS Biology, 10*(1), e1001233.

Stoker, G. (2010). Exploring the promise of experimentation in political science: Micro-foundational insights and policy relevance. *Political Studies, 58*(2), 300–319.

Stoker, G., & John, P. (2009). Design experiments: Engaging policy makers in the search for evidence about what works. *Political Studies, 57*(2), 356–373.

Torgerson, C. J., & Torgerson, D. J. (2013). *Randomised trials in education: An introductory handbook*. London: Educational Endowment Foundation. Retrieved from http://educationendowmentfoundation.org.uk/uploads/pdf/Randomised_trials_in_education_revised.pdf.

Triantafillou, P. (2013). The political implications of performance management and evidence-based poli-
    cymaking. *The American Review of Public Administration*, Online first, April 9, 2013. doi:10.1177/
    0275074013483872.
Tzabar, R. (2013, January 1). *Ben Goldacre's Bad Evidence*. BBC Radio 4.
Weingart, P. (1999). Scientific expertise and political accountability: Paradoxes of science in politics.
    Science and Public Policy, 26(3), 151–161.
Wesselink, A. J., & Hoppe, R. (2011). If post-normal science is the solution, what is the problem? The
    politics of activist science. *Science, Technology and Human Values, 36*(3), 389–412.
Whitty, G. (2013, March 19). *Evidence-informed policy and practice—we should welcome it, but also be
    realistic!* Retrieved June 18, 2013, from https://cerp.aqa.org.uk/perspectives/evidence-informed-
    policy-practice.
Wynne, B. (1992). Uncertainty and environmental learning. *Global Environmental Change, 2*(2), 111–127.
Wynne, B. (2010). When doubt becomes a weapon. *Nature, 466*(7305), 441–442.