Serendipities
Journal for the Sociology and History of the Social Sciences

ARTICLE

When Did Social Experimentation Turn Into Social Science? The History of Randomised Controlled Trials and the Role of Science in Evidence-Based Policymaking

Malte Neuwinger

malte.neuwinger@uni-bielefeld.de

Abstract

Supporters and critics of contemporary evidence-based policymaking (EBPM) widely view randomised controlled trials (RCTs) as an example of social science turned into an instrument of policymaking. Yet, historically, it is more adequate to say that policymaking turned RCTs into social science. Questioning standard views on the history of RCTs, this article argues that for most of the twentieth century, the scientific status of RCTs remained ambiguous—yet this did not stop them from gaining political credibility in social and medical applications. Until recently, experimental approaches were used not because they were regarded as the hallmark of science but because regulators and businesses felt that testing policies and drugs systematically was a sensible thing to do. Debates about the scientific status of RCTs became significant only in the 1980s when some leading social scientists opposed RCTs, viewing them as *un*scientific trial-and-error tinkering. RCTs eventually emerged as a generic instrument of evaluation in the 1990s, when they became regarded as politically credible *and* scientific. Judging from the history of its "gold standard", present-day EBPM's focus on social science instead of politics and business essentially reverses the direction in which it expects evidence to come to policy.

Keywords

randomised controlled trial, social experiment, problem of demarcation, boundary work, history of social science, evidence-based policymaking

INTRODUCTION

Social science is a crucial resource for improving policies and political programs—at least, that is what social scientists and policymakers keep saying. Supposedly, contemporary evidence-based policymaking (EBPM) mines social science for insights and then uses this evidence to develop better policies: more effective ways to help the poor, more sustainable ways to organise the economy, and more sensible ways to tackle acute crises. Whether optimistic or sceptical about achieving these aims,

everyone considers social science the source of potential solutions (Baron 2018; Chupein and Glennerster 2018; Hadorn et al., 2022).

This article examines the history of randomised controlled trials (RCTs)—a kind of social experiment often regarded as the "gold standard" of EBPM—to suggest that the conventional vision sketched above is more recent and less self-explanatory than may first appear. While drawing on the standard repertoire of critical scholarship—asking questions like, What exactly is meant by "science"? What counts as "evidence"? Who is to say what is a "better" policy? (Eyal 2019; Parkhurst 2017; Turner 2013)—the article's main concern is not to question the applicability of RCTs to policy evaluation or to criticise the aims of EBPM. More modestly, it takes on a question asked far less often (Oliver et al., 2014; Stoker and Evans 2016): Why is it necessarily social science that is supposed to make policymaking evidence-based? As it turns out, the answer to this question is far from obvious, especially if by "social science" we mean RCTs.

Over the past twenty years, RCTs have become known as a tool that may "revolutionize social policy" just as they "revolutionized medicine in the twentieth century" (Duflo and Kremer 2005: 117). They are celebrated as a "credibility revolution" in social science (Angrist and Pischke 2010), some of their supporters have recently won a Nobel Prize in economics, and research centres engaged in running and disseminating RCTs define their mission as "ensuring that policy is informed by scientific evidence" (J-PAL 2024). Yet, speaking about RCTs in this way has not always been obvious. As is generally the case historically, experimentation preceded science, and its primary purpose was to guide action rather than improve scientific knowledge (Hansson 2015).

It is widely acknowledged that RCTs in social contexts have been run since the early twentieth century (Baron 2018; Jamison 2019). But what role did social science play? This article argues that, compared to today, researchers and policymakers drew the boundaries between "science" and "non-science" quite differently—and what these boundaries meant varied historically as well. For instance, during the 1960s and 1970s, the success of RCTs was mainly due to bureaucrats and commercial research firms. But unlike today, whether these actors were legitimised through social science was not a standard concern. Later, by the 1980s, leading social scientists, including future Nobel Prize laureate James Heckman, opposed the widespread application of RCTs precisely because they were regarded as *un*scientific trial-and-error tinkering. In their view, creating evidence for policymaking should be left to lesser, non-scientist minds. The idea that social scientists might improve policymaking is not new (Wittrock et al., 1991). Nevertheless, the history of RCTs puts things in perspective, suggesting that the contemporary view of EBPM signifies a shift not only in how we define "rigorous evidence" but also in what we expect from "social science".

What is more, we will see that the dominance of *medical* trials emerged only in the early 1970s, again largely through efforts taken by businesses and government agencies, and became widely regarded as unambiguously scientific only later. Yet when political credibility and scientificity finally came together during the 1990s, "the RCT" was created as the *all-purpose instrument of evaluation* celebrated by contemporary EBPM. For medical and social RCTs, tinkering came first, more serious practical applications followed—and only then, late in the twentieth century, did they come to be widely regarded as "scientific". In today's image of EBPM, the direction of how evidence comes to policy has been essentially reversed, with science assuming a historically unlikely leading position.

The rest of the article proceeds as follows. The next section discusses how the history of RCTs is best understood conceptually. It argues that analysing the "boundary work" among scientists and practitioners helps question whether RCTs were always regarded as part of social science. But at the

same time, putting boundary work centre stage forgets to ask whether RCTs' scientific status always mattered when the technique was applied in twentieth-century policymaking. A perspective modelled on Pierre Bourdieu's sociology of science improves on analyses of boundary drawing, by also considering how categories like "science" and "politics" have changed their meanings over time. The following two sections discuss how the history of RCTs has been written so far and show empirically how taking the shifting boundaries of social science significantly improves our understanding of EBPM. The concluding section considers some objections to the article's main argument, stressing that debates about whether RCTs are part of social science are not just an academic question—the outcome of these debates is one of the main reasons RCTs are seen as the gold standard of EBPM.

RCTS, SCIENCE, AND POLITICAL CREDIBILITY

For the last half-century, one of the workhorses of the sociology of science has been the concept of *boundary work*. Sociologists generally present it as a practical solution to the "problem of demarcation" in the philosophy of science: instead of abstractly debating what activities qualify as "scientific"—something philosophers keep worrying about (Hirvonen and Karisto 2022)—boundary work describes the practical strategies scientists use to position themselves on the side of "science" and others on the side of "non-science" (Gieryn 1983). Crucially, sociologists insist that these practical strategies also have an important practical purpose. Especially in political contexts, whether one's activities are regarded as scientific is an essential source of *credibility* and *legitimacy*. This means scientists and experts will often attempt to preserve their authority and integrity by stabilising, redrawing, or otherwise manipulating the boundaries between science and non-science (Jasanoff 1987).

Historians of RCTs have almost entirely ignored the importance of boundary work. As will be discussed later, they instead draw on a model that remains largely implicit. Reflecting a particular researcher's main interests and expertise, the rise and fall in popularity of RCTs over the twentieth century is usually interpreted either as an outcome of "scientific developments" or "political manoeuvring". In doing so, historical writings consistently associate RCTs with "science"—without checking whether this was how people thought about them at the time. This perspective still guides present discussions. Even contemporary critics usually accept the scientific status of RCTs and focus on other problems, such as the ethical and political implications of experimenting with "human subjects" on a mass scale (Bédécarrats et al., 2020; Fejerskov 2022; de Souza Leão and Eyal 2019).[1]

At first sight, the concept of boundary work seems to enlighten the discussion. Indeed, we will see that zooming in on the instability of the science–politics boundary immediately reveals that unambiguously associating RCTs with science is a very recent novelty. At the same time, it is important to note that most historical writings retain a subtlety that the concept of boundary work tends to conceal. While boundary work makes no assumption about how boundaries are drawn empirically or how they should be drawn epistemically, it makes a strong assumption that the

[1] To be sure, critics have made some scattered angry remarks, according to which the current popularity of RCT is "more a matter of faith than science" (Ravallion 2020: 60), or that "if experimentation were the hallmark of science, there would be Nobel Prizes for alchemy and not for the physics of astronomy" (Pritchett, in Picciotto 2020: 259). But these are not focused analyses of the boundary between science and non-science. Instead, critics question whether RCTs should be regarded as "really scientific", much like the debates in the 1980s we discuss later.

existence of boundaries is universally *significant* and that their significance is *stable* over time. As one researcher expresses the concept's premise:

> It is universally accepted [among all parties involved in discussions of policy-relevant science] that 'science' should not be influenced by politics and that judgements as to what constitutes 'good' science should be left to scientists. All agree, as well, that scientists should not be involved in making 'policy' (Jasanoff 1987: 199).

But while today it seems quite intuitive that credibility and legitimacy flow from boundaries between science and policy—that drawing and maintaining boundaries guards EBPM against turning into "policy-based evidence-making"—the history of RCTs puts a question mark on the transhistorical stability and universal significance of this principle. At various points during the twentieth century, RCTs seemed *credible* enough to base policy decisions on them regardless of their scientific status. The question of whether RCTs were supported by scientists seems to have arisen only later, namely in the 1980s, when academic economists began emphasising the boundary toward non-science explicitly.

Thus, because political credibility, scientific status, and the significance of science–policy boundaries did not move in sync, focusing on boundary work solves one problem and introduces another. It reveals that RCTs became unambiguously associated with science only recently, making it quite anachronistic to speak of historical applications of RCTs as instances of "science influencing policy". Yet the concept is itself anachronistic when assuming that maintaining the boundary between science and policy is equally important in all circumstances.

One way to overcome this conundrum can be gleaned from the work of Pierre Bourdieu (1991, 2004). Similar to the idea of boundary work, Bourdieu insists that social science, politics, or business are never entirely separable. The autonomy of these fields always remains relative, meaning that the question of whether RCTs are scientific may become a significant stake relevant to social actors' credibility. But in addition, Bourdieu argues that everyday categories—again, think of "science", "politics" and "business"—themselves have a history. Accordingly, at different historical moments, such categories are associated with different assumptions and expectations: the boundary between science and politics *may* turn into a stake relevant to RCT proponents' credibility, but this need not happen in all circumstances.

Therefore, it becomes necessary to consider assumptions about "proper" or "normal" boundaries in their historical context, performing an "epistemological break", as Bourdieu liked to say. This makes it possible to newly "construct the object" of analysis—in this case, RCTs—and understand its meaning and significance at a particular point in time (Bourdieu et al., 1991: part 1; Swartz 2013: 21–22).

The following sections will show how Bourdieu's conception (or at least his weariness of others' conceptions) avoids some of the oversights in previous analyses of the history of RCTs, taking seriously the phenomenon of boundary work without necessarily associating claims to political credibility with scientificity, or vice versa. In this process, "the RCT" indeed turns into a new object: one that moves freely between science, politics, and business—but whose credibility emerges from all of them at different times. To demonstrate the value of these arguments, we may begin by considering how the history of RCTs has been written without them.

THE STANDARD PERSPECTIVE ON THE HISTORY OF RCTS

Randomised controlled trials are experimental procedures to compare different options—drugs, policies, technologies, or anything else—about whose causal effects one is not entirely sure. In its simplest form, an RCT randomly assigns people to either a "treatment" group that receives the drug, policy, or whatever it may be, while the other does not. According to the logic of present-day EBPM, if the treatment group shows more promising effects than the control group, the treatment probably "works" in the sense that it likely caused the effect being observed.

The way this procedure has entered policy evaluation is not straightforward. In the United States, interest in RCTs conducted in the social realm—until recently usually called "social experiments" (Hausman and Wise 1985; Riecken and Boruch 1978)—followed a "boom and bust" cycle (Rossi and Wright 1984: 332). At several points, people became excited about experimental methods; a few years later, they lost interest; and yet a few years later, they became excited again. Some scholars have identified three "waves" of RCTs (Jamison 2019), others two (Oakley 2000), and yet others only one (de Souza Leão and Eyal 2019), but the exact number does not matter here.[2] The point is that every serious scholar agrees that the history of RCTs is marked by numerous ups and downs, at least in the United States. (Given that the history of RCTs outside the US remains to be written, it is not the focus of the present discussion).

In reviewing the waves, a *standard perspective* on the history of RCTs emerges. Contemporary historians and RCT experts paint a detailed picture of the ups and downs of experimental studies in social contexts over time. But in framing these ups and downs as a process of "scientific development" or "political manoeuvring", they collectively miss that, for most of the twentieth century, RCTs were not associated with social science in the sense they are today. We will see later that this has significant consequences for how we think about the history of EBPM.

Waves of social experimentation over the twentieth century

Various proto-versions of RCTs have appeared at several points in post-Enlightenment history. Historians of medicine get excited about the eighteenth-century Scottish surgeon James Lind, whose experiments showed that sailors at sea could be prevented from scurvy if they ate enough citrus fruit (Jamison 2019). Similarly, historians of philosophy proudly point out that one of the first people with a good enough grasp of randomisation to use it systematically was the nineteenth-century Pragmatist Charles Sanders Peirce (Hacking 1988). However, these proto-RCTs remained isolated attempts and received only limited attention. Eighteenth and nineteenth-century intellectuals were strikingly united in their lack of interest in taking the experimental procedures common in the natural sciences and applying them to investigations in the social realm (Brown 1997).

The first distinctive wave of social experiments emerged in the 1920s and 1930s. One famous example is the attempt to improve worker productivity at the Hawthorne factory of the Western Electric Company near Chicago, using variations in lighting. Because workers were aware of the ongoing experiments and accordingly adapted their behaviour, we now speak of the "Hawthorne effect" that can complicate the interpretation of experimental studies. Randomised social experimentation continued during World War II (deLeon 1991; Guglielmo 2008). For instance, the US Army's Morale Division—a "psychological warfare" branch inside the military—randomly

[2] One researcher has even identified *four* "waves of evaluation diffusion" (Vedung 2010), but his account extends beyond RCTs in particular.

assigned American soldiers to the film series *Why We Fight*, hoping to learn about its "educational effects". The filmmaking and experimenting were attempts to counter the Germans' successful propaganda efforts, most prominently the movie *Triumph of the Will* (Dehue 2001: 293–294). According to one scholar, American research on the efficacy of propaganda techniques may well have been "the decisive spur to the use of randomized trials to evaluate policy programs" after the war (Oakley 2000: 322).

Another, much larger, wave of RCTs began to roll in 1964, the United States' "War on Poverty". One of the most famous experiments was conducted between 1968 and 1972, a "negative income tax" (NIT) trial that would integrate the social security system into the regular tax system. Because the NIT model posited that people under a certain income threshold would "pay negative taxes"—as in, the government paying *them*—people would effectively be guaranteed an income independent from wage labour (Steensland 2006). While the plan died under later administrations, the NIT trial and its three successors "helped to usher in a brief but heady period of large-scale social experimentation in domestic federal agencies" of the United States government (O'Connor 2001: 190).

Further experiments in policy areas like education, housing, health insurance, and labour market reintegration followed (Hausman and Wise 1985). It was a "golden age of evaluation" (Rossi and Wright 1984: 331). Led by firms including MDRC (then called the Manpower Demonstration Research Corporation), the Institute for Research on Poverty (IRP), Abt Associates, the Urban Institute, and a greatly expanded RAND Corporation, a veritable research industry emerged (Berman 2022: chapter 5).

With the Reagan administration's mission to get rid of the welfare state, the political climate changed during the 1980s. Still, social experiments remained in demand. Indeed, one scholar argues that "quantitative and microeconomic reasoning, with its promise of certainty and rigour and powerful normative foundation, first became really attractive when the content of policies became more controversial" (Jann 1991: 124). The Social Security Act of 1962 allowed states to apply for waivers. If states pledged to design "innovative" (i.e., in this case, conservative) policy reforms, standard welfare law would not apply. However, to get the waivers approved, the federal government required states to test their proposed reforms in experimental studies (Harvey et al., 2000). One labour economist notes that this arrangement made RCTs "almost mandatory" during the late 1980s and early 1990s (Moffitt 2004: 520). This de facto obligation ended in 1996 when Bill Clinton announced his plan to "end welfare as we know it". However, as we know today, experimental methods only took a few years to regain traction.

"Science" and "politics" as drivers of RCTs?

Why were RCTs popular at one point in time and less popular at another? Why were they so popular in the United States rather than elsewhere? Following the standard perspective, historians of social science and experts on RCTs suggest that it must have something to do with political priorities and the relative relevance of social science in different countries and at different times—how developed the social sciences were and how much attention politicians paid to the insights of social scientists (Backhouse and Fontaine 2010; Levitt and List 2009; O'Connor 2001). But while the idea that some combination of "science" and "politics" is responsible for the varying popularity of RCTs over time seems natural, the intuitiveness of these categories tends to insulate them from scrutiny. As we will see now, almost all historical narratives dubiously depict RCTs as a technique that was "scientific" all along, even though their status was ambiguous at best.

The United States was the first country to develop the kind of professionalised social science we know today. Early on, American academics established the conviction that "first and foremost social science was *science*, not philosophy, or social reform, or history" (Bulmer 1991: 164, emphasis in original). They explicitly imagined themselves "as detached, nonjudgmental observers rather than as helping neighbours or political allies" (O'Connor 2001: 47). With some exceptions, the rest of Europe lagged behind these efforts (Backhouse and Fontaine 2010; Wittrock et al., 1991). To historians of social science, it follows that developments in social science explain the fondness of Anglo-American countries for experimental methods. Having the most policy-focused social sciences and the most social science-oriented political tradition gave them a head-start.

RCT experts have built on this general characterisation, filling the gaps with empirical detail. RCT enthusiasts, in particular, generally name *scientific* factors for the first wave of experimental trials, such as breakthroughs in statistics. Along with his earlier publications, Ronald Fisher's 1935 book *The Design of Experiments* is credited with being "a main catalyst for the actual use of randomization in controlled experiments" in various contexts (Levitt and List 2009: 3). Similarly, RCT experts commonly trace the second wave to a famous 1966 textbook by Donald T. Campbell and Julian Stanley, *Experimental and Quasi-Experimental Designs for Research* (Rossi and Wright 1984: 334). Invariably, RCT experts explain the increasing application of experimental methods in policy research through methodological advances in social science. Some, like Campbell and Stanley themselves, even attributed the *end* of each wave to the same factor. In their view, scientific advances simply had not been significant enough (Campbell and Stanley 1967; Rossi 1987).

More historically-minded scholars—who are usually more critical of using experimental methods in policy evaluation—do not dismiss science as a significant force of social change, nor do they doubt the scientificity of RCTs. Still, they stress the importance of *politics*. They argue that social problems increasingly came to be seen as technical issues with technical solutions—which rendered scientific breakthroughs applicable to politics in the first place. During the War on Poverty, social scientists could count on "hitherto unheard-of amounts of financial support and ready access to policy makers" (deLeon 1991: 91). For instance, the first NIT experiment, conducted in New Jersey, was based on an initial idea by graduate student Heather Ross. Worth at least $US5 million at the time (over $US30 million in today's dollars), it was "perhaps one of the most expensive doctoral theses in economics" (Levitt and List 2009: 5).

What is more, the NIT experiments were enabled by a striking political U-turn. While early efforts of the War on Poverty had been primarily focused on community programs aimed at organising poor people politically, a media scandal involving several Chicago street gangs led to the installation of former RAND Corporation staff in the relevant government agencies. By October 1965, these new personnel had substituted community organising for a more conservative policy of income maintenance and job creation (O'Connor 2001: 171–178).

During the 1980s, the innovation waivers discussed previously kept RCTs alive, but other political factors stopped them from spreading further. Because experimental interventions were less generously funded, they became smaller and more selective. As the welfare state came under fire, doubts about the contributions of social science also became prominent. In the words of a group of distinguished historians, the Thatcher and Reagan governments defunded the social sciences because they saw them, and "in particular the non-economic approaches, as biased towards the welfare state and towards the cause of the political opponent" (Wittrock et al., 1991: 53). Subsequent research tended to focus on checking existing programs' cost-effectiveness, hoping to guard the state against high expenditures.

It is obvious that scientific developments and political manoeuvring are indispensable for understanding the history of RCTs. The historical facts are not in question. At the same time, historians of social science themselves agree that the conceptual distinction that (however innocently) structures their narratives is hard to maintain. For instance, they occasionally note that because social scientists made increasing "efforts to tackle social problems" after World War II, the "traditional domains of the social sciences were redefined" in post-war societies (Backhouse and Fontaine 2010: 184). Another scholar writes that "the greater methodological rigour of American, as opposed to British, research" might be due to the "position of social scientists in the two societies and their self-conception" (Bulmer 1991: 159).

Still, historians and RCT experts generally treat experimental techniques as unambiguously "scientific" and explain their popularity over time through the clearly distinct variables "science" and "politics." The following section emphasises the fact that these variables are far more permeable than often assumed.

RCTS BEYOND SCIENCE AND POLITICS

The power of theory is greatest when its existence remains implicit. The implicit theory inherent in the standard perspective on the history of RCTs sounds simple enough and broadly "fits the facts" it purports to explain. What other than scientific and political factors would be responsible for RCTs' ups and downs over time?

This section describes how a view modelled on Bourdieu's sociology of science changes the picture. As indicated previously, this perspective stresses that the boundaries between science and politics are dynamic rather than static. It also recognises that social scientists rarely operate entirely autonomously from "heteronomous" considerations outside the area demarcated as science, meaning that those least favoured by the epistemic norms of the scientific orthodoxy will be most inclined to appeal to "non-scientific" considerations (Bourdieu 2004: 85–87). Finally, beyond exclusively analysing the permeability of the boundaries between science and politics, Bourdieu insists on putting the categories of "science" and "politics" themselves in historical context. Whether an activity is viewed as scientific *can* become a relevant stake related to political credibility—but whether and when it does cannot be assumed in advance (Bourdieu et al., 1991: part 1; Swartz 2013).

These considerations suggest that the standard perspective's assumptions are too neat in some cases and actively misleading in others. As we will see, one example is that the "golden age" of RCTs during the 1960s and 1970s was due to the rising political influence of economists, as some scholars have correctly observed (de Souza Leão and Eyal 2019: 392–394), but not necessarily due to "social scientists". At the time, RCTs became attractive because economists successfully established counterfactual thinking as a new criterion of political credibility. Yet unlike in today's EBPM, whether RCTs (or their supporters) were regarded as scientific was not a primary concern.

A second example is that, in the 1980s, influential social scientists considered RCTs as "too political" to be part of "science" proper—making them turn against both RCTs and against intervening politically. When governments and research firms began to endorse RCTs, social scientists turned the very meaning of "social science" into a subject of boundary work. Finally, a third example of the limitation of the standard perspective is that *clinical* trials were not considered scientific for much of the twentieth century either. Only when the evidence-based medicine movement finally merged clinical trials' political credibility with scientific ambition, a generic instrument of evaluation called "the RCT" was born.

Credible RCTs without social science, social science without RCTs

The contemporary success of RCTs relies on the fact that a large group of social actors—social scientists, governments, non-governmental organisations, private research firms, and philanthropic foundations—supports them. They all agree on RCTs' credibility in scientific *and* applied contexts (Neuwinger 2024). Two debates in the area of labour market research, where RCTs have had the greatest success over the longest timeframe, illustrate that this view is relatively novel historically. The first debate, from the 1970s, shows how a new *criterion of credibility*—the "criterion of net impact"—was established in the US government, sowing the seeds for large-scale RCTs in social policy. The second, from the 1980s, features two groups of economists, one advocating a detached and the other a more applied type of analysis, arguing about *whether RCTs were proper social science*. Together, these debates show how the political credibility and scientific status of RCTs emerged—but largely independently of each other.

Do labour market policies, such as programs teaching unemployed people labour market-relevant skills, improve their earnings? Before the 1970s, answering this question was easy. Program administrators simply checked whether a person's earnings were higher after participating in their program than in previous employment. But if you think like an economist, this is wrong. And because the number of economists employed in government had continuously increased since World War II, they would not let administrators get away with common sense (Berman 2022). After all, economists said, people's earnings were likely to rise simply because of favourable macroeconomic conditions or mere luck. The assumed causal relation between the training program and higher earnings was spurious. A more adequate criterion of program success, economists argued, was based on "the counterfactual": whether a person's earnings would also have been higher if they had *not* been part of the program. This is what they called the *criterion of "net impact"*. According to economists, counterfactual reasoning was a better guide to assessing impact than the factual experience of program administrators (Breslau 1997a: 882–884).

Why would government officials want to evaluate their programs against hypothetical situations dreamed up by economists? The short answer is that they were essentially forced to do so. Fierce discussions ensued. Eventually, however, the Assistant Secretary for Policy Evaluation and Research (ASPER), the Office of Management and Budget (OMB), and the Department of Labor (DOL) used their combined executive power to mandate the use of the net impact criterion (Berman 2022: 108–110; Breslau 1997a: 889–891). By the late 1970s, the terms of debate had changed in favour of the economists. It was not that everyone had changed their minds, but that economists had used their increasing influence in government circles to establish their criteria of good argument inside its leading institutions. The government had become guided by an "economic style of reasoning" (Berman 2022: 5–6). It was now thinking counterfactually. Today, we may describe this new standard of reasoning as more scientific than the old one. But notably, it was not established by maintaining a solid boundary between "science" and "policy" (or by blurring it) but rather by ignoring its relevance.

This first debate does not yet have much to do with RCTs, only with establishing the criterion that made them *politically credible*. This is where the second debate begins. After economists had established the net impact criterion in government thinking, the state hired numerous research firms to run RCTs. Yet, at some point, this led to the question of whether the approach favoured by the government was actually *scientific*. It was only then that boundary work between science and politics became a significant principle for the historical trajectory of RCTs.

By the 1980s, the problem was that people calling themselves "social scientists"—those employed at universities—would not touch RCTs. To them, interdisciplinary fields like program evaluation or the "policy sciences" remained largely suspect (Brunner 1982). Professional evaluators and academic social scientists were "living in worlds with different cultures and senses of what was at stake" (Gueron and Rolston 2013: 271). The situation resembled what the small group of people who *did* support RCT *and* were employed by universities termed a "Catch-22":

> ...applied social research [primarily meaning RCTs] is relatively less prestigious inside the university; therefore, university researchers are not attracted to it, particularly as a full-time career line. And yet the reason why applied research is not highly regarded in the first place is that not much of it goes on in the university! (Rossi et al., 1978: 181).

While evaluation companies conducted sophisticated experiments, their efforts were not considered a part of legitimate social science. Why? One important reason was that a group around James Heckman, an econometrician who would win the Nobel Prize in 2000, held that sophisticated statistical modelling made RCTs largely unnecessary (Heckman 1979, 2020). Heckman's followers were opposed by a more diverse group of economists employed by research firms and academics frequently contracting with the government, who argued that proper experimental tests proved the assumptions of econometric models unrealistic (Breslau 1998: 87–88, see Fraker and Maynard 1987; LaLonde 1986). Overall, econometricians attempted to protect the discipline of social science from trial-and-error tinkering, while social experimenters attacked the academic orthodoxy from the scientific margins of applied research. To keep social science "pure", most university-based researchers could not approve of RCTs and their "contract shop epistemology" (Breslau 1997b).

The disdain social experimenters and social scientists held for each other is amusing enough to quote at length. Even when being interviewed in the 1990s, one of them still fumed:

> [People doing RCTs at commercial research firms] have absolutely no respect for data. They have absolutely no respect for knowledge. They're beltway bandits, pure and simple, just a bullshit operation. They have absolute contempt for anything. What they're good at is sort of greasing the palms of Labour Department bureaucrats with low IQs. And that's basically what they have got in the pocket right now (quoted in Breslau 1997b: 377–378).

To the people facing this kind of critique, things probably would not have appeared quite as amusing. Judith Gueron, long-time president of MDRC, the research firm mentioned previously, recalls that at the time, she was "fortunately [...] not aware of the depth of vitriol or of the personal disdain some academics felt toward what they called 'contract shops'" (Gueron and Rolston 2013: 270).

In any case, while the debate partly mapped on the two parties' concerns for money and prestige, it was also an implicit fight about the *primary concerns of social science*. Academic economists held the immediate problems of government administration at arm's length, while social experimenters insisted on helping governments solve concrete problems. Because social experimenters' RCTs remained "directly action-guiding" rather than "epistemic" (Hansson 2015)—trying to find out whether things work in practice rather than how they work in theory—social scientists refused to regard them as science and remained sceptical of them.

Overall, the two debates show that RCTs are a far less stable object than assumed in contemporary EBPM or the standard historical perspective. Their credibility and scientific status had to be established through active struggles among various actors. And while boundary work is part of these struggles, the story is further complicated by the fact that the political credibility and scientific status

of RCTs changed independently of each other. However, one additional piece of a better understanding of the history of RCTs is how their credibility and scientific status finally came together. As I discuss now, this merging involved developments in medicine.

Medical trials, medical science, and the emergence of "the RCT" as a generic instrument of evaluation

Much rhetoric on the "gold standard" status of RCTs in present-day EBPM depends on the intuitiveness of an analogy: that running RCTs in social contexts is roughly equivalent to running RCTs in medical contexts. Following Nobelists Duflo and Kremer's (2005: 117) exclamation that experimental studies may "revolutionize social policy" just as they "revolutionized medicine in the twentieth century", medicine is commonly used as a scientific success story other fields should seek to emulate (see Leigh 2018: chapter 2; Manzi 2012: chapter 7).

Yet such an argument from analogy would not have been plausible for much of the twentieth century. As we will see, this is because medicine was itself a latecomer to RCTs, and it was hardly "science" that convinced the majority of medical professionals. As with social policy RCTs, the scientific status and political credibility of medical RCTs emerged largely independently. Experimental approaches were popularised and mandated on a large scale by commercial firms and the state during the 1960s and 1970s. Only later did medical experimentation become recognised as unambiguously scientific and turned into a politically credible blueprint for RCTs in the social realm. This finally turned "the RCT" into a *generic instrument of evaluation*, endorsed in scientific and political contexts and applicable to a large variety of problems.

Clinical trials are "newer than the automobile or the airplane, and about the same age as the color television or the electronic computer" (Manzi 2012: 78). While this fact is seldom made explicit, it is hardly controversial. Similar to the econometricians of the 1970s discussed previously, before the 1950s, doctors widely deemed randomised experiments unnecessary and unethical. Because medical professionals were confident about their knowledge of how to treat illnesses, randomly subjecting people to treatments "known" not to work seemed outrageous (Bothwell et al., 2020; Marks 2000: chapter 1). Yet, little by little, simplified versions of statistical insights (which had also initiated the first wave of social experiments in the 1920s and 1930s) gained ground among the medical establishment. While battles over whether medicine could or should be a science kept raging, by the 1960s, statisticians had convinced significant numbers inside the medical profession that experimental methods were worth considering, at least in principle (Marks 2000: chapter 5).

Yet the primary motivation for the widespread adoption of RCTs did not emerge through science, as understood then or today. Instead, it was due to capitalism and the determination of states to protect their citizens from the adverse effects of an all-too-free market. As the pharmaceutical industry expanded, regulatory agencies had to deal with large numbers of often unproven and sometimes dangerous drugs. Perhaps the most significant push came as a reaction to the thalidomide crisis of the late 1950s and early 1960s. First marketed by the West German company Grünenthal as *Contergan*, it soon turned out that the drug many women took against morning sickness during pregnancy led to massive increases in stillbirths and babies being born without arms or legs (Bothwell et al., 2020).

These events prompted the US Food and Drug Administration (FDA) to write some surprisingly stringent regulations. By 1970 (but not earlier), all pharmaceutical products had to undergo a sophisticated process involving randomised and double-blinded experimental studies. If this did not

lead to official approval, selling them was illegal. The new standard of demanding RCTs was soon adopted by all regulatory agencies of the industrialised world (Bothwell et al., 2020; Carpenter 2014: 238–245). By now, a global industry of contract research organisations (CRO) has taken over data collection and analysis (Petryna 2009). What we think of today as the "scientific standards" of medical trials could just as well be described as "industry standards" created and imposed by state power.[3]

So then, why do we regard these (very sensible) standards as so unambiguously scientific that they can work as a blueprint for social RCTs in contemporary EBPM? Given the early resistance among doctors and the involvement of regulatory agencies, one might guess quite the opposite. Indeed, the most significant case of boundary work to render medicine genuinely scientific (and insist on this status) succeeded only in the early 1990s with the emergence of "evidence-based medicine", or EBM (Daly 2005; Evidence-Based Medicine Working Group 1992; Sacket et al., 1996). Until the 1980s, medicine had been an "impure science" whose standards and scientific aspirations were constantly challenged (Epstein 1996). Yet, piggybacking on the new experimental orthodoxy in pharmaceutical research, EBM advocates like Cochrane began to clarify the evidence base of clinical practice and promoted systematic reviews, meta-analyses, and experimental evidence as the most scientific approach to medicine (Daly 2005).

One interesting example of how advocates worked the boundaries comes from the 1992 article that first popularised EBM in academic circles. Explicitly drawing on Thomas Kuhn's theory of science, the article referred to its project as "a new paradigm for medical practice" (Evidence-Based Medicine Working Group 1992: 2420). The dominance of clinical trials and meta-analyses in the assessment of new drugs, the article claimed, was what Kuhn had meant when he talked about paradigms as "ways of looking at the world that define both the problems that can legitimately be addressed and the range of admissible evidence that may bear on their solution" (ibid.).

EBM thus explicitly directed attention away from physiological mechanisms and common sense and toward advanced statistics and experimental methods. Much like Kuhn suggested for the development of science, it popularised these convictions in guidelines, textbooks, courses and journals. EBM also constantly referred to a selection of core exemplars, such as experimental tests of tuberculosis and polio in the 1940s and 1950s (Solomon 2011: 454–457). According to Kuhn, "the member of a mature scientific community is, like the typical character of Orwell's 1984, the victim of a history rewritten by the powers that be" (Kuhn 1970: 167). As the powers of EBM established their new paradigm of proper medical science, they rewrote history in their own sense.

Overall, the history of medical and social RCTs is almost analogous. As with social trials, clinical trials first became popular and politically credible in practical applications, outside the realm then understood as science. Like social scientists did at the time, doctors long rejected experimental methods, and it required tremendous regulatory pressure to make the pharmaceutical industry change course. Like social RCTs, clinical trials have been turned into the epitome of science only recently, through the successful boundary work of the EBM movement. Indeed, this movement was so successful that, in effect, it established RCTs as a *generic instrument of evaluation*—one that was credible politically *and* scientifically, and that could, in principle, be used in the medical *and* social realm.

[3] Policy researchers are used to concepts like "regulatory science" to describe the hybrid of scientific and regulatory knowledge common in regulatory agencies, but present RCT debates are not so subtle.

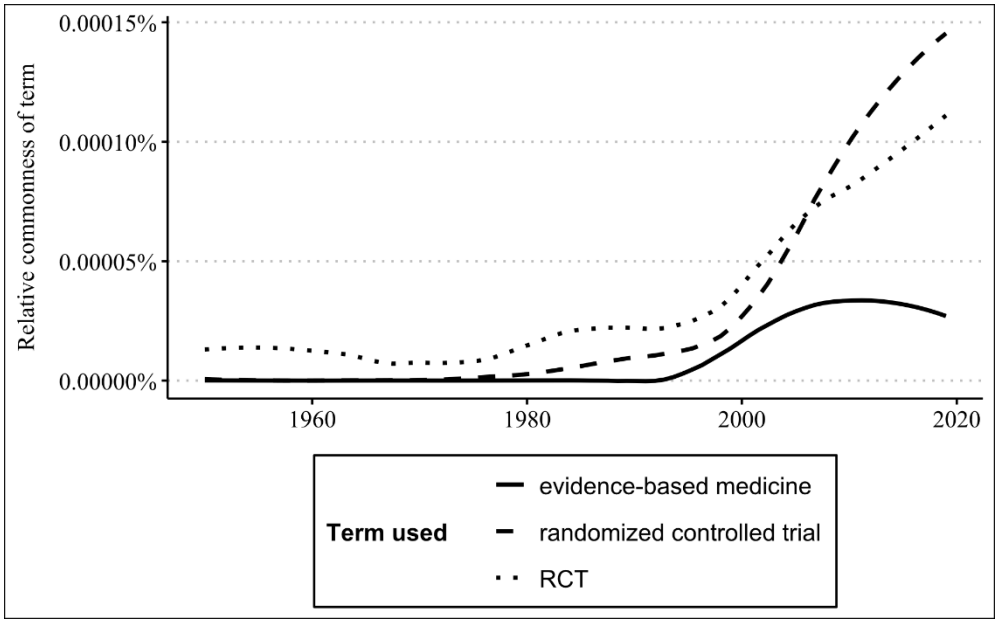

Figure 1 Mentions of "randomised controlled trials" and "evidence-based medicine" in general and scholarly literature, 1950 – 2019. Data: Google Books Ngram Viewer, 2019 corpus.

In a sense, one may even say that the rise of EBM created "the RCT" as such. As shown in Figure 1, the 1980s and 1990s were the time when the term "randomised controlled trial" first became common usage, making the hitherto familiar "social experiment" sound as suspicious as it sounds today. After medicine provided a plausible analogy for political credibility and unambiguous scientificity, RCT supporters could begin building the coalitions between players from science, business and politics that make today's EBPM movement so strong (Neuwinger 2024). In one sense, running a *social experiment* in the 1960s or an *RCT* in the 2020s is precisely the same activity. But in another sense, the history of experimental trials is genuinely discontinuous, beginning anew when "the RCT" emerged as an all-purpose evaluation tool of unambiguous scientific standing.

CONCLUSION

The present wave of RCTs was preceded by many previous ones. Historically, the contemporary "credibility revolution" in social science and EBPM builds on at least a hundred years of randomised social experimentation. But while in today's EBPM discussions, political credibility goes together with claims of social scientific rigor, these two components of the credibility revolution remained separate for much of the twentieth century.

Observing the history of RCTs from a contemporary perspective thus tends to obscure the fact that the credibility of experimental studies was established by governments and commercial research firms, whose actions we associate with social science only in retrospect. It also obscures the fact that some leading social scientists used to oppose RCTs during the 1980s—and that they did so not because they were "anti-science" but because they were "pro-science". Finally, the contemporary perspective obscures the fact that clinical trials (and medicine more generally) emerged as a generally accepted role model for EBPM only by the early 1990s, when scientific and political credibility came together with support from the evidence-based medicine movement. The notion that policy should be guided by social science is flattering to social scientists. But the history of the present "gold standard" of EBPM suggests that this notion is far less obvious than it may first seem—the outcome of subtle shifts in the meaning and purpose of social science.

Before concluding, it is worth considering two objections to these arguments. First, historically-minded people may point out that the reconstruction attempted here is no news to anyone. After all, isn't it common knowledge that progress and innovation have very often emerged from practical, economic, or regulatory endeavours rather than "independent science"? Isn't this what the history of statistics teaches us (Porter 1986), and couldn't the whole history of experimentation—not only of RCTs—be told according to the catchphrase that "tinkering and technology preceded science"? (Hansson 2015). These historical remarks are quite apt, and they indicate, as historically-minded people are fond of putting it, "what RCTs are a case of". However, the response to these questions is that most historical writings have, in fact, *not* treated RCTs as a case of "technology before science" or of boundary shifts in what we expect from "social science". If the argument is not particularly original, it is all the more notable that RCTs are usually discussed as an obvious example of "science influencing policy".

This leads directly to the second objection. More scientifically-minded people may ask why it matters at all whether RCTs are regarded as "scientific" or "technological" or anything else. After all, these are only words, and what matters is not words but whether RCTs can actually improve policymaking. Hasn't the "problem of demarcation" between science and non-science been shown to be quite unimportant long ago, a mere "*machines de guerre* in a polemical battle" about nothing in particular? (Laudan 1983: 119, emphasis in original). Again, this objection is understandable. Yet it only emphasises the practical importance of boundary work. Dismissing matters of demarcation as sophistry—insisting that merely verbal debates should be ignored—underestimates the way sophistries about "science" and "non-science" might be precisely the kind of *socially significant sophistry* that keeps contemporary EBPM debates going. While their track record of improving policymaking is mixed at best, RCTs are nevertheless seen as the "gold standard" for achieving improvements in the future. Maintaining this hope largely depends on present-day RCT supporters' success in presenting their favourite tool as unambiguously scientific.

REFERENCES

Angrist, Joshua D. and Jörn-Steffen Pischke (2010) The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics, *Journal of Economic Perspectives* 24(2): 3–30.

Backhouse, Roger E. and Philippe Fontaine (2010) Toward a History of the Social Sciences, in Roger E. Backhouse and Philippe Fontaine (Eds.), *The History of the Social Sciences Since 1945*, Cambridge, UK: Cambridge University Press, 184–233.

Baron, Jon (2018) A Brief History of Evidence-Based Policy, *The ANNALS of the American Academy of Political and Social Science* 678(1): 40–50. https://doi.org/10.1177/0002716218763128.

Bédécarrats, Florent, Isabelle Guérin, and François Roubaud (eds.) (2020) *Randomized Control Trials in the Field of Development: A Critical Perspective*, Oxford: Oxford University Press.

Berman, Elizabeth Popp (2022) *Thinking Like an Economist: How Efficiency Replaced Equality in U.S. Public Policy*, Princeton, N.J.: Princeton University Press.

Bothwell, Laura E., Wen-Hua Kuo, David S. Jones, and Scott H. Podolsky (2020) Social and Scientific History of Randomized Controlled Trials, in Steven Piantadosi and Curtis L. Meinert

(Eds.), *Principles and Practice of Clinical Trials*, Cham: Springer International Publishing, 1–17.

Bourdieu, Pierre (1991) The peculiar history of scientific reason, *Sociological Forum* 6(1): 3–26. https://doi.org/10.1007/BF01112725.

Bourdieu, Pierre (2004) *Science of Science and Reflexivity*, Cambridge, UK: Polity Press.

Bourdieu, Pierre, Jean-Claude Chamboredon, and Jean-Claude Passeron (1991) *The Craft of Sociology: Epistemological Preliminaries*, Berlin: De Gruyter.

Breslau, Daniel (1997a) The Political Power of Research Methods: Knowledge Regimes in U. S. Labor-Market Policy, *Theory and Society* 26(6): 869–902.

Breslau, Daniel (1997b) Contract Shop Epistemology: Credibility and Problem Construction in Applied Social Science, *Social Studies of Science* 27: 363–394.

Breslau, Daniel (1998) *In Search of the Unequivocal. The Political Economy of Measurement in U.S. Labor Market Policy*, Westport, CT: Praeger.

Brown, Robert (1997) The delayed birth of social experiments, *History of the Human Sciences* 10(2): 1–21.

Brunner, Ronald D. (1982) The policy sciences as science, *Policy Sciences* 15(2): 115–135.

Bulmer, Martin (1991) National contexts for the development of social-policy research: British and American research on poverty and social welfare compared, in *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*, Cambridge, UK: Cambridge University Press, 148–167.

Campbell, Donald T. and Julian C. Stanley (1967) *Experimental and quasi-experimental designs for research*, Boston: Houghton Mifflin Company.

Carpenter, Daniel (2014) *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*, Princeton, N.J.: Princeton University Press.

Chupein, Thomas and Rachel Glennerster (2018) Evidence-Informed Policy from an International Perspective, *The ANNALS of the American Academy of Political and Social Science* 678(1): 62–70.

Daly, Jeanne (2005) *Evidence-based medicine and the search for a science of clinical care*, Berkeley: University of California Press.

Dehue, Trudy (2001) Establishing the Experimenting Society: The Historical Origin of Social Experimentation According to the Randomized Controlled Design, *The American Journal of Psychology* 114(2): 283–302.

deLeon, Peter (1991) Political events and the policy sciences, in Björn Wagner, Carol Hirschon Weiss, Björn Wittrock, and Hellmut Wollmann (Eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*, Cambridge, UK: Cambridge University Press, 86–109.

Duflo, Esther and Michael Kremer (2005) Use of Randomization in the Evaluation of Development Effectiveness, in George Keith Pitman, Osvaldo N. Feinstein, and Gregory K. Ingram (Eds.), *Evaluating Development Effectiveness*, Piscataway, NJ: Transaction Publishers, 93–120.

Epstein, Steven (1996) *Impure science: AIDS, activism, and the politics of knowledge*, Berkeley: University of California Press.

Evidence-Based Medicine Working Group (1992) Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine, *JAMA* 268(17): 2420–2425.

Eyal, Gil (2019) *The Crisis of Expertise*, Cambridge, MA: Polity Press.

Fejerskov, Adam (2022) *The Global Lab: Inequality, Technology, and the Experimental Movement*, Oxford: Oxford University Press.

Fisher, Ronald (1935) *The Design of Experiments*. Edinburgh: Oliver & Boyd Ltd.

Fraker, Thomas and Rebecca Maynard (1987) The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs, *The Journal of Human Resources* 22(2): 194–227.

Gieryn, Thomas F. (1983) Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists, *American Sociological Review* 48(6): 781–795.

Gueron, Judith M. and Howard Rolston (2013) *Fighting for Reliable Evidence*, New York: Russell Sage Foundation.

Guglielmo, Mark (2008) The Contribution of Economists to Military Intelligence During World War II, *The Journal of Economic History* 68(1): 109–150.

Hacking, Ian (1988) Telepathy: Origins of Randomization in Experimental Design, *Isis* 79(3): 427–451. https://doi.org/10.1086/354775.

Hadorn, Susanne, Fritz Sager, Céline Mavrot, Anna Malandrino, and Jörn Ege (2022) Evidence-Based Policymaking in Times of Acute Crisis: Comparing the Use of Scientific Knowledge in Germany, Switzerland, and Italy, *Politische Vierteljahresschrift* 63(2): 359–382. https://doi.org/10.1007/s11615-022-00382-x.

Hansson, Sven Ove (2015) Experiments Before Science. What Science Learned from Technological Experiments, in *The Role of Technology in Science: Philosophical Perspectives*, Dordrecht: Springer Netherlands, 81–112.

Harvey, Carol, Michael J. Camasso, and Radha Jagannathan (2000) Evaluating Welfare Reform Waivers Under Section 1115, *Journal of Economic Perspectives* 14(4): 165–188. https://doi.org/10.1257/jep.14.4.165.

Hausman, Jerry A. and David A. Wise (1985) *Social Experimentation*, Chicago: University of Chicago Press.

Heckman, James J. (1979) Sample Selection Bias as a Specification Error, *Econometrica* 47(1): 153–161.

Heckman, James J. (2020) Epilogue: Randomization and Social Policy Evaluation Revisited, in Florent Bédécarrats, Isabelle Guerin, and François Roubaud (Eds.), *Randomized Control Trials in the Field of Development*, Oxford: Oxford University Press, 304–330.

Hirvonen, Ilmari and Janne Karisto (2022) Demarcation without Dogmas, *Theoria* 88(3): 701–720. https://doi.org/10.1111/theo.12395.

Jamison, Julian C. (2019) The Entry of Randomized Assignment into the Social Sciences, *Journal of Causal Inference* 7(1): 1–16. https://doi.org/10.1515/jci-2017-0025.

Jann, Werner (1991) From policy analysis to political management? An outside look at public policy training in the United States, in Peter Wagner, Carol Hirschon Weiss, Björn Wittrock, Hellmut Wollman (Eds.), *Social Sciences and Modern States: National Experiences and Theoretical Crossroads*, Cambridge: Cambridge University Press, 110–130.

Jasanoff, Sheila (1987) Contested Boundaries in Policy-Relevant Science, *Social Studies of Science* 17(1): 195–230.

J-PAL (2024) The Abdul Latif Jameel Poverty Action Lab [online]. https://www.povertyactionlab.org/ [Last accessed: 22 January 2024].

Kuhn, Thomas S. (1970) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

LaLonde, Robert J. (1986) Evaluating the Econometric Evaluations of Training Programs with Experimental Data, *The American Economic Review* 76(4): 604–620.

Laudan, L. (1983) The demise of the demarcation problem, in R.S. Cohen and L. Laudan (Eds.), *Physics, Philosophy and Psychoanalysis*, Dordrecht: Springer Netherlands, 111–127.

Leigh, Andrew (2018) *Randomistas: How Radical Researchers Are Changing Our World*, New Haven: Yale University Press.

Levitt, Steven D. and John A. List (2009) Field Experiments in Economics: The Past, the Present, and the Future, *European Economic Review* 53(1): 1–18.

Manzi, Jim (2012) *Uncontrolled: The surprising payoff of trial-and-error for business, politics, and society*, New York: Basic Books.

Marks, Harry M. (2000) *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900-1990*, Cambridge, UK: Cambridge University Press.

Moffitt, Robert A. (2004) The Role of Randomized Field Trials in Social Science Research. A Perspective From Evaluations of Reforms of Social Welfare Programs, *American Behavioral Scientist* 47(5): 506–540.

Neuwinger, Malte (2024) What makes randomized controlled trials so successful—for now? Or, on the consonances, compromises, and contradictions of a global interstitial field, *Theory and Society* 53(5): 1213–1244.

Oakley, Ann (2000) A Historical Perspective on the Use of Randomized Trials in Social Science Settings, *Crime & Delinquency* 46(3): 315–329.

O'Connor, Alice (2001) *Poverty Knowledge: Social Science, Social Policy, and the Poor in Twentieth-Century US History*, Princeton, N.J: Princeton University Press.

Oliver, Kathryn, Theo Lorenc, and Simon Innvær (2014) New directions in evidence-based policy research: a critical analysis of the literature, *Health Research Policy and Systems* 12(1): 34.

Parkhurst, Justin O. (2017) *The politics of evidence: from evidence-based policy to the good governance of evidence*, London: Routledge.

Petryna, Adriana (2009) *When Experiments Travel. Clinical Trials and the Global Search for Human Subjects*, Princeton, N.J.: Princeton University Press.

Picciotto, Robert (2020) Are the 'Randomistas' Evaluators?, in *Randomized Control Trials in the Field of Development*, Oxford University Press, 256–279.

Porter, Theodore M. (1986) *The Rise of Statistical Thinking, 1820–1900*, Princeton, N.J.: Princeton University Press.

Ravallion, Martin (2020) Should the Randomistas (Continue to) Rule?, in Florent Bédécarrats, Isabelle Guerin, and Francois Roubaud (Eds.), *Randomized Control Trials in the Field of Development*, Oxford: Oxford University Press, 47–78.

Riecken, H.W. and R.F. Boruch (1978) Social Experiments, *Annual Review of Sociology* 4: 511–532.

Rossi, Peter H. (1987) The Iron Law of Evaluation and Other Metallic Rules, *Research in Social Problems and Public Policy* 4: 3–20.

Rossi, Peter H. and James D. Wright (1984) Evaluation Research: An Assessment, *Annual Review of Sociology* 10: 331–352.

Rossi, Peter H., James D. Wright, and Sonia R. Wright (1978) The Theory and Practice of Applied Social Research, *Evaluation Quarterly* 2(2): 171–191.

Sacket, David, William Rosenberg, Jamuir Gray, Robert Brian Haynes, and Scott Richardson (1996) Evidence based medicine: what it is and what it isn't, *British Journal of Medicine* 312: 71–72.

Solomon, Miriam (2011) Just a paradigm: evidence-based medicine in epistemological context, *European Journal for Philosophy of Science* 1(3): 451–466.

de Souza Leão, Luciana and Gil Eyal (2019) The rise of randomized controlled trials (RCTs) in international development in historical perspective, *Theory and Society* 48(3): 383–418.

Steensland, Brian (2006) Cultural Categories and the American Welfare State: The Case of Guaranteed Income Policy, *American Journal of Sociology* 111(5): 1273–1326.

Stoker, Gerry and Mark Evans (2016) Evidence-Based Policy Making and Social Science, in Gerry Stoker and Mark Evans (Eds.), *Evidence-Based Policy Making in the Social Sciences*, Bristol: Policy Press, 15–28.

Swartz, David L. (2013) Metaprinciples for Sociological Research in a Bourdieusian Perspective, in Philip S. Gorski (Ed.), *Bourdieu and Historical Analysis*, Durham: Duke University Press, 19–35.

Turner, Stephen P. (2013) *The Politics of Expertise*, New York: Routledge.

Vedung, Evert (2010) Four Waves of Evaluation Diffusion, *Evaluation* 16(3): 263–277.

Wittrock, Björn, Peter Wagner, and Hellmut Wollmann (1991) Social science and the modern state: policy knowledge and political institutions in Western Europe and the United States, in *Social Sciences and Modern States: National Experiences and Theoretical Crossroads*, Cambridge, UK: Cambridge University Press, 28–85.

ACKNOWLEDGEMENTS

Research for this article was conducted during my time with the Research Training Group "World Politics" at Bielefeld University. I thank the Special Issue guest editors for their incisive comments, which have greatly improved the manuscript – at least the sections I was not too stubborn to rewrite.

AUTHOR BIOGRAPHY

Malte Neuwinger is a sociologist based at Bielefeld University, Germany.