13. Methods development in evidence synthesis: a dialogue between science and society

James Thomas

INTRODUCTION

This chapter is about the science of evidence synthesis: the way that academics bring together knowledge from across multiple studies into a whole, to present the state of current understanding about a given area. While scientists have been doing this for centuries in the form of literature reviews, both the methods used to combine findings from multiple studies, and the purpose for which this is done have undergone significant changes in the past 30–40 years. First, the methods used to conduct traditional literature reviews have been found wanting, and more formal, systematic, methods have been developed. Second, the systematic (literature) review has become one of the primary formal routes through which research findings are mediated beyond the academy to inform decision-making in policy and practice settings. The close dialogue between decision-makers and systematic review authors has asked uncomfortable questions of the academic community about how research can be synthesised, and how valid and robust conclusions can be drawn from such syntheses. It is part of a wider story of a shifting power dynamic between universities and society in which society is increasingly expecting operationalisable 'solutions' from research efforts that can directly improve societal and individual outcomes (Collini, 2012; Nowotny et al., 1994). While this may be problematic in some settings, there can be benefits for both sides: the pressure from members of society on the academy to produce usable research can be a positive, in that it both demonstrates that the knowledge produced by research is considered useful and the intercourse between science and society can help the science to be conducted in socially relevant ways.

This chapter starts by giving an introduction to the way that evidence synthesis began and what its methods were before moving on to consider how increased engagement with users of reviews has changed the way that reviews are done. It concludes by taking stock of where we are currently and considers some of the main challenges yet to be addressed. The focus is on systematic reviews that aim to inform the selection and implementation of interventions. Many other research questions are answered by systematic reviews – including diagnostic, prognostic and epidemiological – but it is in the context of interventions that most systematic reviews are conducted, and the main issues about the development of methods are most clearly identified in this review type.

THE RECENT GENESIS OF EVIDENCE SYNTHESIS

Contemporary accounts of evidence synthesis and systematic reviews often start with Gene Glass, an American educational psychologist and statistician working at the University of Colorado Boulder. During the 1970s, he conducted a 'meta-analysis' of studies in the area of psychotherapy in order to settle a long-running dispute in the field as to whether or not psychotherapy 'worked'. To answer this question, Glass compiled a dataset of psychotherapy studies and, using what we now know as 'meta-analysis' to combine their results, demonstrated a beneficial effect for psychotherapy (Glass, 2015). This work was highly significant because, as well as advancing knowledge in the field of psychotherapy, it established an important principle: in order to speak with authority about the state of knowledge in an area, it was necessary to analyse all relevant studies. It also saw the birth of a new statistical method: meta-analysis, for combining the results from multiple studies. (While the term 'meta-analysis' has sometimes been used to mean the whole endeavour of identifying studies, assessing their relevance and reliability, and summarising their findings, it is more generally used now simply to mean the statistical aggregation of findings from multiple studies, and it is this latter interpretation that is used throughout this chapter.)

The work of Gene Glass was extended into education in the 1980s, and then, most significantly, into health. During the 1970s and 1980s there had been a growing realisation that medicine was not 'evidence-based', and that many decisions concerning patient care were taken in ignorance of the current state of the evidence. What became the 'evidence-based medicine' (EBM) movement was born out of this realisation and the writings and actions of David Sacket and colleagues at McMaster University in the late 1980s and early 1990s. It was concerned with 'the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients'; and focused heavily on the training and professional development of heath care professionals (Sackett et al., 1996). One of the means through which EBM might be achieved was by using systematic reviews, and by helpful coincidence the 'Cochrane Collaboration' was formed around the same time in response to a challenge from Archie Cochrane made more than a decade before (1979): 'It is surely a great criticism of our profession that we have not organised a critical summary, by speciality and subspeciality, adapted periodically, of all randomised controlled trials.'

It is helpful to bear in mind, however, that while the Cochrane Collaboration and EBM are often considered as sharing the same goals, they in fact had quite separate beginnings, and have slightly different missions. While the EBM movement was concerned with changing clinical practice through clinician training and engagement, the Cochrane Collaboration's objective has been to supply the world with reliable summaries of research. The Collaboration set about its task in the 1990s and early 2000s to identify randomised trials and pull their results together in systematic reviews – using meta-analysis as the primary method of synthesis. It was largely a volunteer effort, with individual reviews often led by specialists in their field who wanted to keep on top of the emerging evidence base. The collaboration publishes its reviews in the 'Cochrane Library', available first on floppy disks and later online. The 1990s also saw the Collaboration publish software to support its reviewers in the form of the review authoring tool, RevMan. This tool enables reviewers to write their reviews in a highly structured form, and to conduct the statistical meta-analytic component of the review. This aided publication, but did, however, limit the range of analytical techniques available to the pairwise comparison of two treatments (or a single treatment versus placebo) in randomised trials that all made the same comparison. Demand for systematic reviews and for the Collaboration's outputs has grown over the years, and it now attracts significant funding from research funders and income derived from the Cochrane Library, which now contains more than 8,000 systematic reviews. In parallel with the development of the Cochrane Collaboration, wider interest in systematic reviews has grown quickly. From the early 1990s, when hardly any were published, tens of thousands of systematic reviews are now being published each year (shown in Figure 13.1 below[1]).

The history of the Cochrane Collaboration and the development of methods for evidence synthesis are instructive, because they reveal an endeavour that is strongly embedded in the research community, with systematic reviews published in a similar way to research articles in an academic journal. (Indeed, while many countries have subscriptions to the Cochrane Library, and it is available open access in low- and middle-income countries, it is published by an academic publisher and requires a subscription to access.) Thus, while systematic reviews of research often aim to inform decisions outside universities, they began life as a university-based endeavour, rather than something driven by a demand for systematic reviews from outside the academy.

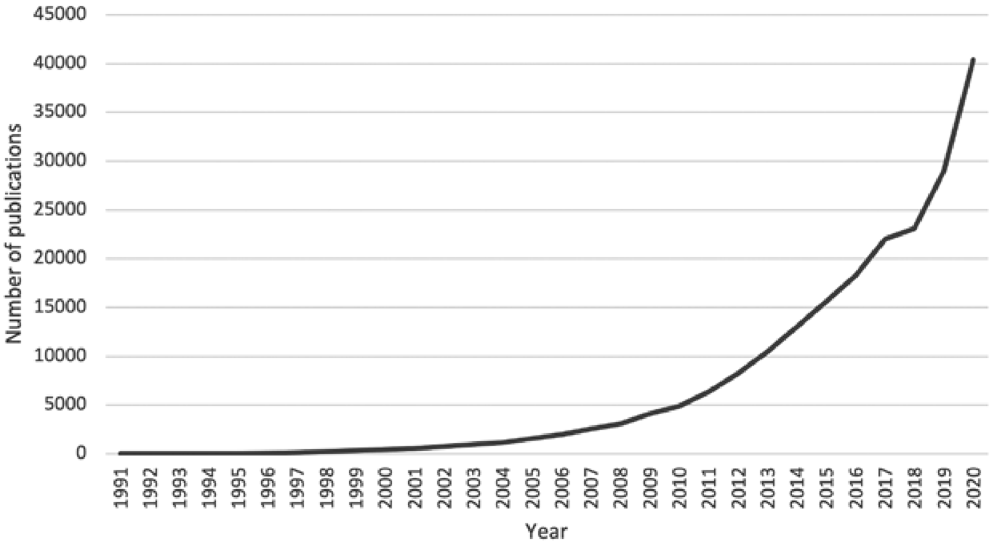

Figure 13.1 Growth in publications of systematic reviews

EARLY METHODS OF EVIDENCE SYNTHESIS

Having overviewed the early history of systematic reviews and evidence synthesis, this section examines some of the principles and priorities that underpinned early methods of evidence synthesis. EBM had (and has) a 'hierarchy' of evidence for addressing questions of effectiveness, ordered in terms of the confidence that can be placed in their results (Greenhalgh, 1997). As assessing the effectiveness of an intervention (or treatment) involves making a causal connection between an intervention and the outcome(s) observed, research designs that offer the most confidence in their causal claims are prioritised. Regarded as least reliable are case reports and cross-sectional surveys, because of their relative weakness in establishing that the intervention *caused* the outcome, rather than another factor. While not without their critics, randomised (controlled) trials have been generally held to provide the strongest evidence for a causal claim linking an intervention with an observed outcome, and these are placed near the top of the hierarchy (Kleijnen et al., 1997). The only step 'above' a randomised trial is a systematic review of randomised trials, since individual studies can still be atypical, and a systematic review containing multiple trials can estimate intervention effects more accurately and confidently than individual studies.

Early systematic reviews fitted within the paradigm of knowledge prioritised by EBM, aiming to provide the strongest causally robust evidence about the effects of interventions. Not only did this involve a heavy focus on the randomised trial study design, but also a concern with other threats to the reliability of the evidence produced in systematic reviews. Bias might be introduced into systematic reviews in two ways: arising from the way that the review itself was done; and from bias in the findings of primary studies included. Concern about the first aspect has led to strict methodological expectations about the way that systematic reviews are conducted (Higgins et al., 2021). They are expected to follow a protocol which is written and published before the review work starts, specifying in detail the searching and analytical methods that will be used. Using a protocol aims to minimise bias that might creep into the review process when researchers have seen the results of the research that will be included, offering it 'procedural objectivity' (Reiss and Sprenger, 2014). Reviews also need to ensure that they are free from selection bias, and so require extensive, sensitive searching for relevant studies over multiple electronic databases and aim to minimise the impact of publication bias (where some studies are systematically withheld from publication (Easterbrook et al., 1991). This caution extends to the way that reviews are done and the way that mistakes in their conduct are minimised: so many tasks in a systematic review are done (and checked) by more than one person. Finally, the studies included in systematic reviews are appraised for 'risks of bias', which examine threats to their internal validity (Higgins et al., 2011).

The early meta-analyses conducted by Gene Glass involved the statistical aggregation of research. This focus on statistics continued during the 1980s with the highly influential work of Hedges and Olkin (Hedges and Olkin, 1985) and, when the Cochrane Collaboration was formed in the 1990s, it was again assumed that systematic reviews would synthesise study findings using statistical meta-analysis. When combining studies in a meta-analysis to estimate a combined effect size estimate (and confidence intervals), the underpinning assumption is that the studies are estimating the same underlying effect and differ from one another by little more than random / sampling error. This carries with it the corollary assumption that the studies in the review are evaluating the same intervention with the same population (and measuring outcomes using the same tools). While this demand is often met when clinical treatments are concerned, it is less likely to be the case when the focus moves to, for example, public health or education, where exact replications of interventions among the same population are relatively rare.

Early systematic review methods then, were developed within universities in the context of EBM and inherited its focus on questions of effectiveness and its preference for assessing causality using randomised trials. The priority was in generating valid evidence to demonstrate that specified interventions did, or did not, result in a specified outcome. This epistemology is heavily influenced by medical epidemiology, counterfactual models of causation, and scientific empiricism (Illari and Russo, 2014). In this paradigm, if the counterfactual evidence is sufficiently strong, it is not always necessary to know *how* and *why* a given intervention has the effect it does; it is enough to know that the application of a given intervention will result in the desired outcome.

INCREASING ENGAGEMENT WITH USERS OF REVIEWS

During the 1990s, the idea of using systematic reviews of research to inform decision-making grew quickly (Young et al., 2002). They became an important part of the UK's National Institute of Clinical Excellence[2] to guide the development of clinical guidelines. They were also the focus of a UK Department of Education and Skills project to promote the use of evidence in education (Oakley et al., 2005). Their use spread beyond the UK to organisations such as the WHO and the US Agency for Healthcare Research and Quality. In 2000, the Campbell Collaboration was founded, a sister organisation to Cochrane, which aimed to support systematic reviews in social science, including education, crime and justice, and social welfare. Since then, systematic reviews have been undertaken in numerous fields including the environment and international development, social care, business, software development, and pre-clinical research.

Systematic reviews were part of wider moves to see policy and practice decisions informed by up-to-date research with a greater expectation from society for universities to produce actionable research findings, and for them to have a positive 'impact' on society. This expectation has become increasingly explicit with, for example, greater weight given to 'impact case studies' in the UK's Research Excellence Framework, and 'pathways to impact' a formal part of the application process for research funding. Cochrane and other organisations had developed mechanisms for including the perspectives of stakeholders (or 'consumers', as Cochrane calls them) in reviews, but during the early years of systematic review production, these processes were mostly concerned with involving stakeholders in reviews that were conceived and led by academics. This has been termed a 'push' model of research utilisation, where academics identify a problem in wider society that they consider research can help with; the evidence synthesis is conducted, and research is 'pushed' to its intended audience (Weiss, 1979). While many reviews have employed this method successfully – and their findings have been used by their intended audiences – societal expectations of research have grown, and with this, 'external' stakeholders have exerted increasing influence over not only the questions asked in reviews, but the way that they are done, and the methods they use. Timeliness has been a perennial challenge for systematic reviews, as finding and synthesising all the research on a given topic can take more than a year to complete. Users of research often need systematic review findings more quickly than this and have pushed systematic reviews to be conducted more quickly. The questions addressed by systematic reviews have also come under scrutiny, since the priorities of academics may not be the same as those who are wanting, for example, to develop practitioner guidance. Cochrane has been under pressure over both the timeliness and relevance of its reviews by its main UK funder over many years. In a sign of just how much influence a body 'external' to academia wants to have in academic processes, the National Institute for Health Research issued a stark warning in 2021 that the writing was now 'on the wall': either Cochrane ensured that its reviews address the right questions at the right time, or its funding would be cut.[3] Without seemingly giving Cochrane much time to respond, a few months later the NIHR announced that all infrastructure funding for Cochrane would cease. As Helga Nowotny (2004) asks: 'Society is moving into a position where it is increasingly able to communicate its wishes, desires and fears to Science. What happens then to science as result of this reverse communication?'

METHODOLOGICAL IMPLICATIONS FOR REVIEWS

The changes wrought by society's increased involvement in systematic reviews have been extremely far-reaching. One of the challenges has certainly been about timeliness, but this is mainly about processes and how researchers can be more efficient. More fundamental have been the challenges relating to the questions asked. For systematic reviews that are concerned with interventions, the changes have ranged from advancing statistical methods to developing approaches that conflict with the values that systematic reviewers have traditionally held dear. The next few paragraphs overview some of the main developments, reflecting on the type of knowledge generated.

One relatively recent development has been in expanding the scope of a meta-analysis so that it provides a more formal comparison of different interventions. The way that early systematic reviews addressed intervention effectiveness – in terms of comparing a single intervention with another or placebo – have often been found wanting. These reviews are often characterised as 'what works?' reviews, but the means of answering them are more akin to asking a 'does this work?' question. Decision-makers often need to choose from a list of possible options, and simply assessing pairwise comparisons of intervention x versus no intervention does not speak directly to this decision. Addressing this gap has been the recent rise in interest of 'mixed treatment comparisons', or 'network meta-analysis'. In this method, many possible interventions are compared with one another, providing a 'ranked list' of intervention effectiveness. The value of such approaches has been apparent during the COVID-19 pandemic, with regularly updated network meta-analyses able to inform decision-makers which out of more than 20 treatments were most effective, and where more research was required (Siemieniuk et al., 2020). Network meta-analyses require more work than conventional systematic reviews because they include so many comparisons, and they also require more advanced statistical techniques. They do not, however, fundamentally challenge the type of evidence produced: its account of causation is still empirical, dependent on valid counterfactuals, and can be produced using relatively conventional meta-analytic techniques.

More challenging to this type of knowledge are questions for which strong counterfactual evidence is not available. For example, in education and other areas of social research, exact replications of intervention evaluations in the same population are rare. More common are 'variations on a theme', where the same broad approach is tested in different ways, and where different components are combined to make up 'multicomponent' interventions. This scenario challenges the above conceptualisation of a meta-analysis, where the assumption is that the same intervention is being evaluated in much the same population, because there is no single intervention being evaluated. While a 'random effects' meta-analysis aims to account for the additional variation in outcomes that this situation introduces, the problem is rooted in the nature of the knowledge being produced and cannot be addressed by statistical methods alone. If one combines the results from slightly different interventions in slightly different populations and obtains an overall pooled effect size estimate, what does this estimate actually represent? And how does a decision-maker use it? Which of the variations in approach, or combination of intervention components, is the one that they should choose? And how do they decide which might be most appropriate in a given population? There are, of course, statistical techniques that enable review authors to conduct sub-group analyses and try to identify which components are most important for particular populations, but two factors significantly limit their utility. First, the number of ways that interventions and populations differ from one another often vastly exceeds the number of studies available for analysis; there are often insufficient data to produce valid results. Second, most statistical techniques necessarily model linear relationships between independent variables, and research into intervention complexity has shown that this is not a safe assumption. Here, phase changes (where something needs to reach a given threshold before an effect is seen), feedback loops and non-linear effects are expected and, moreover, some important determinants of intervention success, such as socioeconomic status, are simply not amenable to randomisation. The empiricist counterfactual approach to systematic reviews cannot produce evidence with the kind of confidence in its causal claims that is possible when considering medical treatments (see, e.g., Kelly, 2018; Petticrew et al., 2012; Petticrew, 2015; Petticrew et al., 2019; Thomas, et al., 2014).

Systematic review authors have responded to these challenges in a number of ways. First, there has been conceptual and methodological work to extend existing statistical methods to address increasing complexity (Higgins et al., 2019; O'Mara-Eves and Thomas, 2016). These developments have mostly been used in practice to support subgroup analysis and meta-regression that can clarify what might drive differences between study outcomes. Importantly, though, while a subgroup analysis might successfully 'explain' why some studies report larger effects than others, this comparison is essentially 'observational' in nature, even when the comparison is between randomised trials. This reduction in counterfactual security means that the approach is often described in methods texts as being suitable for 'hypothesis generation' only, rather than it being on the same level as the rest of the review's results. Authors find themselves in the position of being 'damned if you do, damned if you don't' when attempting to address research questions, such as those involving equity and the exploration of inequalities, for which randomised trial evidence does not exist. On the one hand, there is a need to address these questions because they are asked by users of reviews. On the other, traditional statistical methods say that these kinds of subgroup analyses should not be used to draw firm conclusions for practice, because the approach does not meet expected standards for establishing cause and effect (Petticrew et al., 2012).

Second, attempts have been made to strengthen the validity of causal observational claims with relevant theory. Some approaches to this involve the development of 'logic models' which aim to express, diagrammatically, the context within which an intervention has been introduced, and which might explain variation in its impact (Anderson et al., 2011; Rogers, 2000). Logic models can operate at different levels of the review and might encapsulate the system of interest (e.g., education system and drivers of particular policies and practices), or specific processes that the review aims to elucidate (e.g., the causal pathways that might link a new pedagogical approach with education outcomes) (Rehfuess et al., 2017). Methods for developing logic models are still in their infancy, and a mixture of theoretical literature and practice knowledge are often employed. More formally, theory can be used to construct subgroup analyses in 'mixed methods' systematic reviews. Here, a synthesis of the qualitative research often precedes a subgroup analysis of the results of randomised trials, though sometimes the qualitative evidence is used after the meta-analysis to explain its findings (Pluye and Hong, 2014; Thomas et al., 2004; van Grootel et al., 2017). One novel approach that was originally developed by Ragin in the context of political science is Qualitative Comparative Analysis (QCA) (Ragin, 2008). QCA aims to transcend the qualitative / quantitative divide and combine the iteration and focus on theory that is usually the province of qualitative research, with formal quantitative methods for combining statistical results. It combines mechanistic, regularity (where inference is drawn from patterns in the data) and counterfactual accounts of causation but is relatively rarely used at present.

Third, entirely new approaches to synthesising study findings have been developed. Perhaps the most popular example here is Realist Review, developed by Pawson, who was dissatisfied with the epistemic limitations of conventional systematic reviews (Pawson, 2002). Realist reviews are based on scientific and critical realism, and so understand causation as 'generative'; that is, generated by mechanisms that are often unobservable. The aim of a realist review is to identify and understand how these mechanisms are triggered in different contexts to result in observed outcomes. The emphasis on a mechanistic approach to causation puts this approach in stark contrast to the priority given to counterfactuals in conventional reviews.

Fourth, the exponentially growing research literature, along with the demand for increasingly complex reviews, is making the task of keeping on top of the evidence base more and more difficult. The 'rapid review', which may be more limited in scope or have fewer quality assurance processes than a full systematic review, has been advanced as a potential solution in some situations, though defining exactly what a 'rapid review' is, is not straightforward (Tricco et al., 2015). Another possible solution has come from new technologies, including machine learning. There are numerous papers that support the idea of using artificial intelligence (AI) in systematic reviews (Tsafnat et al., 2013), but aside from a few specific case studies, such as a machine learning model that can identify randomised trials for Cochrane (Thomas et al., 2021), this new technology has not seen widespread adoption. This might be considered surprising, given the considerable workload burden a systematic review involves, and the fact that numerous (admittedly small-scale) evaluations find consistently that machine learning can reduce the amount of manual work needed (O'Mara-Eves et al., 2015). There are two possible reasons for this. The first reason is that the predominant culture among those doing and using systematic reviews is inherently cautious. Reviewers will go to great lengths to ensure that they identify all relevant studies and that their reviews are reliable, and they are naturally cautious about new approaches that might threaten the reliability of established methods (Arno et al., 2021; O'Connor et al., 2019). Second, the tools themselves may be at a fairly early stage in their evolution and may not yet be sufficiently accurate for use in decision-making workflows that affect people's lives. Evaluations of many tools claim substantial workload savings, but at the cost of losing 5 per cent or more relevant records. This is often considered to be unacceptably high in a context where existing search methods are expected to lose no more than about 1 per cent of the records that should be identified. There are also concerns about possible bias in some of the current state-of-the-art tools, where societal, racial, gender and other biases are reproduced in the machine learning models (Bender et al., 2021).

Finally, there are of course many other questions to ask about interventions, such as whether a given intervention is acceptable or appropriate in a given context; and how a particular intervention approach might be understood. Here, qualitative research is usually synthesised using techniques that are 'borrowed' from primary qualitative analysis. The first method to be developed was meta-ethnography, by Noblit and Hare (1988). Though the first publication on this method was in 1988, it was in a very specific context of use and the method was only 'discovered' by the wider systematic review community, and adopted for use beyond the synthesis of ethnographies, at the start of the 21st century. In a qualitative evidence synthesis (QES), the papers and reports of primary studies are often treated as 'data', and an attempt made to understand not only the perspectives of the participants in the original studies, but also those of the study authors. The objective of a QES is to develop theory which might be used to understand how an intervention might be received and understood by participants. On the whole, they are not used to assess the impact of interventions per se and so, after some initial challenges when they were introduced, are not seen to be as threatening to the conventional systematic review paradigm as some of the other approaches listed above.

IMPLICATIONS

The previous section has highlighted some of the methodological developments in systematic reviews over the past couple of decades. Overall, they have been driven by a need to address a wider range of research questions and in different topic areas than conventional approaches as pioneered in the 1990s. The increased interest in reviews beyond the academy – and indeed, in valuing the knowledge they contain – lies behind many of the developments. One of the major challenges to this sphere of science in response to society's increased involvement is around epistemology: the fact that the questions asked by society cannot be answered within the epistemic security that randomised trials and counterfactual accounts of causality can offer. This echoes findings from Nowotny and colleagues in their book *Re-thinking Science* where they conclude that increased societal involvement in science will necessarily bring changes in the way that science is conducted (Nowotny, Scott, and Gibbons, 2001). They introduced the concepts of 'mode 1' and 'mode 2' knowledge, where 'mode 1' describes more traditional academic research, and 'mode 2', the newer knowledge, is often generated in the context of its application outside the academy. Table 13.1 below highlights some of the differences between the two modes of knowledge, summarising how mode 2 knowledge has different priorities, being concerned more with relevance and social robustness than mode 1, and how it demands much more context specificity – including in terms of time – than its mode 1 counterpart. While this can be a helpful heuristic to understand some of the recent dynamics shaping knowledge production, the 'mode 1 mode 2' construction has not met with universal approval and has been criticised as being over-simplistic, limited in global applicability, and possibly simply being a reformulation of the long-debated division between 'basic' and 'applied' research (Etzkowitz and Leydesdorff, 2000).

Table 13.1 Mode 1 versus mode 2 knowledge[4]

Mode 1	Mode 2
Closed-isolated	Open-interconnected
Objective/reliable	Socially robust
Decontextualised	Context dependent
Restricted to scientific communities	Not restricted to scientific communities
Disciplinary	Inter- and trans-disciplinary
Longevity	Time-dependent

This has not always been a comfortable journey for academics – and the destination may not yet be in sight. University academics have been troubled by the 'interference' of society in problems that should rightly be the preserve of the academy; the consequential impact of this on the methods used, and how to manage the difficult balance of preserving scientific reliability whilst supporting (some degree of) social robustness. To give just a few examples, one of the papers 'celebrating' the 20th anniversary of the Cochrane Collaboration gives insights into the discussions and arguments that accompanied the introduction of qualitative evidence into the Cochrane Library (Hannes et al., 2013). The group met resistance from within the collaboration, summarised in a section entitled 'mixing progress with rejection', which describes how the inclusion of qualitative evidence was resisted by some, and at times this resistance resulted in the group being unable to move forward. Another example of resistance comes from a quite different quarter, with one methodologist in the field of QES expressing their horror at the 'metasynthetic madness' they perceived in the wide use of methods for QES developed to produce findings for policy and practice (Thorne, 2017). What is interesting about this paper is the strong argument mounted in favour of a particular 'purist' approach to qualitative research, without any concern about the use of research beyond the academy: while the target of their ire was the type of evidence synthesis conducted, it might equally have been directed at the involvement of society in shaping the methods in the first place. In another discipline, when the EPPI-Centre was commissioned to support systematic reviews in education in the early 2000s, numerous 'resistance' articles were published criticising perceived governmental control of the research agenda and the misuse of a positivist 'medical model' in education (Oakley, 2006). 'Advocacy of systematic reviews, and of evidence-based practice, are closely associated with moves towards increased central control over educational research…' argued one critic, Hammersley, who questioned 'whether researchers should be cooperating with this erosion of academic autonomy, given the deleterious effects it is likely to have on the future quality of research' (Hammersley, 2001). Hammersley's defence of academic autonomy was taken to the extreme in a postmodern polemic by MacLure who argued that 'systematic review is just one part, I suggest, of a pervasive attempt to reconfigure and regulate professional and academic practices and identities by acting on the very words that people speak and write'(MacLure, 2005). So dangerous were systematic reviews and 'the evidence movement' she said, that 'the time has come for concerted opposition'.

As the previous paragraph has shown, the dialogue between science and society over the synthesis of research to inform decisions has been heated at times. There has thus far been little 'concerted opposition', possibly because this is part of a wider societal change, with more porous walls between universities, industry and society, and the expectation that knowledge can be found at the click of a button. The problems faced by systematic reviewers are not yet solved either. The systematic review 'catechism' is likely to undergo further reshaping, as there are still no obvious solutions to that fundamental conundrum of addressing the range and depth of questions asked whilst maintaining the epistemic security sought by many academics (Petticrew, 2015). Nowotny's question about what happens to science as a result of societal involvement is as relevant now as when it was asked more than a decade ago. Some of the answers have been outlined above in the development of methods that grapple with complexity and bring together different types of research to provide a more holistic picture; it would appear we have some way to go, however, before the journey is complete.

NOTES

1. This figure is indicative, and the numbers likely to be a considerable under-estimate. It was generated from a simple search of 'systematic review' OR 'meta-analysis' in Microsoft Academic during April 2021.
2. Later combined with the Health Development Agency and now called the National Institute for Health and Care Excellence.

3. Ken Stein, Director of the NIHR Evidence Synthesis Programme, speaking at ‘Virtually Cochrane’, April 2021: https://www.youtube.com/watch?v=ukr7B39pyio accessed 23 September 2023.
4. Based on: Kuklinski et al. (2011), Understanding Media, Today. *International Journal of McLuhan Studies.*

REFERENCES

Anderson, L. M., Petticrew, M., Rehfuess, E., Armstrong, R., Ueffing, E., Baker, P., Francis, D. and Tugwell, P. (2011). Using logic models to capture complexity in systematic reviews. *Research Synthesis Methods*, *2*(1), 33–42. https://doi.org/10.1002/jrsm.32.

Arno, A. D., Elliott, J., Wallace, B., Turner, T., and Thomas, J. (2021). *The views of health zguideline developers on the use of automation in health evidence synthesis*, *10*(16), 1–10. https://doi.org/10.21203/rs.3.rs-23742/v1.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big? FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Vol. 1). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Cochrane, A. (1979). 1931–1971: a critical review, with particular reference to the medical profession. In *Medicines for the year 2000*.

Collini, S. (2012). *What are Universities for?* Penguin.

Easterbrook, P. J., Gopalan, R., Berlin, J. A. and Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, *337*(8746), 867–872. https://doi.org/10.1016/0140-6736(91)90201-Y.

Etzkowitz, H. and Leydesdorff, L. (2000). The dynamics of innovation: From National Systems and ‘mode 2’ to a Triple Helix of university-industry-government relations. *Research Policy*, *29*(2), 111. https://doi.org/10.1016/S0048-7333(99)00055-4.

Glass, G. V. (2015). Meta-analysis at middle age: A personal history. *Research Synthesis Methods*, *6*(3), 221–231. https://doi.org/10.1002/jrsm.1133.

Greenhalgh, T. (1997). How to read a paper: Getting your bearings (deciding what the paper is about). *British Medical Journal*, *315*(7102), 243–246. https://doi.org/10.1136/bmj.315.7102.243.

Hammersley, M. (2001). On ‘Systematic’ Reviews of Research Literatures: A ‘narrative’ response to Evans and Benefield. *British Educational Research Journal*, *27*(5), 543–554. https://doi.org/10.1080/01411920120095726.

Hannes, K., Booth, A., Harris, J. and Noyes, J. (2013). Celebrating methodological challenges and changes: reflecting on the emergence and importance of the role of qualitative evidence in Cochrane reviews. *Systematic Reviews*, *2*, 84. https://doi.org/10.1186/2046-4053-2-84.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press.

Higgins, J. P. T., Lasserson, T., Chandler, J., Tovey, D., Thomas, J., Flemyng, E. and Churchill, R. (2021). *Methodological Expectations of Cochrane Intervention Reviews (MECIR). Standards for the conduct and reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning, conduct and reporting of updates*. London.

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L. and Sterne, J. A. C. (2011). The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ*, *343*(18 Oct), d5928–d5928. https://doi.org/10.1136/bmj.d5928.

Higgins, J. P. T., López-López, J. A., Becker, B. J., Davies, S. R., Dawson, S., Grimshaw, J. M., McGuinness, L. A., Moore, T. H. M., Rehfuess, E. A., Thomas, J. and Caldwell, D. M. (2019). Synthesising quantitative evidence in systematic reviews of complex health interventions. *BMJ Global Health*, *4*(Suppl 1), e000858. https://doi.org/10.1136/bmjgh-2018-000858.

Illari, P. and Russo, F. (2014). *Causality: Philosophical Theory Meets Scientific Practice*. Oxford: Oxford University Press.

Kelly, M. P. (2018). The need for a rationalist turn in evidence-based medicine. *Journal of Evaluation in Clinical Practice*, (May), 1–8. https://doi.org/10.1111/jep.12974.

Kleijnen, J., Gøtzsche, R., Kunz, A., Oxman, A. and Chalmers, I. (1997). So what’s so special about randomisation? In A. Maynard and I. Chalmers (eds), *Non-random Reflections on Health Services*

Research: On the 25 Anniversary of Archie Cochrane's Effectiveness and Efficiency (pp. 93–106). London: BMJ Books.
MacLure, M. (2005). Clarity bordering on stupidity: where's the quality in systematic review? *Journal of Education Policy*, *20*(4), 393–416. https://doi.org/10.1080/0268093050013180.
Noblit, G. and Hare, D. (1988). *Meta-ethnography*. Sage.
Nowotny, H. (2004). The potential of transdisciplinarity. *H. Dunin-Woyseth, H. and M. Nielsen, Discussing Transdisciplinarity: Making Professions and the New Mode of Knowledge Production, the Nordic Reader, Oslo School of Architecture, Oslo, Norway*, 10–19. https://doi.org/10.1007/978-3-0348-8419-8_7.
Nowotny, H., Schwartzman, S., Scott, P., and Trow, M. A. (1994). *The New Production of Knowledge*. Thousand Oaks: SAGE Publications.
Nowotny, H., Scott, P. and Gibbons, M. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Wiley-Blackwell.
O'Connor, A. M., Tsafnat, G., Thomas, J., Glasziou, P., Gilbert, S. B. and Hutton, B. (2019). A question of trust: Can we build an evidence base to gain trust in systematic review automation technologies? *Systematic Reviews*, *8*(1), 1–8. https://doi.org/10.1186/s13643-019-1062-0.
O'Mara-Eves, A. and Thomas, J. (2016). Ongoing developments in meta-analytic and quantitative synthesis methods: Broadening the types of research questions that can be addressed. *Review of Education*, *4*(1), 5–27. https://doi.org/10.1002/rev3.3062.
O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M. and Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews*, *4*(1), 5. https://doi.org/10.1186/2046-4053-4-5.
Oakley, A., Gough, D., Oliver, S. and Thomas, J. (2005). The politics of evidence and methodology: lessons from the EPPI-Centre as a case-study. *Evidence and Policy*, *1*(1), 5–31.
Oakley, A. (2006). Resistances to 'new' technologies of evaluation: Education research in the UK as a case study. *Evidence and Policy*, *2*(1), 63–87. https://doi.org/10.1332/174426406775249741.
Pawson, R. (2002). Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation*, *8*(3), 340–358. https://doi.org/10.1177/135638902401462448.
Petticrew, M., Tugwell, P., Kristjansson, E., Oliver, S., Ueffing, E. and Welch, V. (2012). Damned if you do, damned if you don't: subgroup analysis and equity. *Journal of Epidemiology & Community Health*, *66*(1), 95–98. https://doi.org/10.1136/jech.2010.121095.
Petticrew, M. (2015). Time to rethink the systematic review catechism? Moving from 'what works' to 'what happens.' *Systematic Reviews*, *4*(1), 1–6. https://doi.org/10.1186/s13643-015-0027-1.
Petticrew, M., Knai, C., Thomas, J., Rehfuess, E. A., Noyes, J., Gerhardus, A. … McGill, E. (2019). Implications of a complexity perspective for systematic reviews and guideline development in health decision making. *BMJ Global Health*, *4*(Suppl 1), e000899. https://doi.org/10.1136/bmjgh-2018-000899.
Pluye, P., and Hong, Q. N. (2014). Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. *Annual Review of Public Health*, *35*, 29–45. https://doi.org/10.1146/annurev-publhealth-032013-182440.
Ragin, C. (2008). *Redisigning Social Inquiry: Fuzzy Sets and Beyond. University of Chicago Press*.
Rehfuess, E. A., Booth, A., Brereton, L., Burns, J., Gerhardus, A., Mozygemba, K., Oortwijn, W., Pfadenhauer, L. M., Tummers, M., van der Wilt, G.-J. and Rohwer, A. (2017). Towards a taxonomy of logic models in systematic reviews and health technology assessments: A priori, staged, and iterative approaches. *Research Synthesis Methods*, *9*(1), 13–24. https://doi.org/10.1002/jrsm.1254.
Reiss, J., and Sprenger, J. (2014). Scientific Objectivity. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2020 Edition)*. Retrieved 23 September 2023 from https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/.
Rogers, P. J. (2000). Causal models in program theory evaluation. *New Directions for Evaluation*, *87*(87), 47–55. https://doi.org/10.1002/ev.1181.
Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., and Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71.
Siemieniuk, R. A., Bartoszko, J. J., Ge, L., Zeraatkar, D., Izcovich, A., Pardo-Hernandez, H., Rochwerg, B., Lamontagne, F., Han, M. A., Kum, E., Liu, Q., Agarwal, A., Agoritsas, T., Alexander, P., Chu,

D. K., Couban, R., Darzi, A., Devji, T., Fang, B. and Fang, C. (2020). Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ (Clinical Research Ed.)*, *370*, m2980. https://doi.org/10.1136/bmj.m2980.

Thomas, J., Harden, A., Oakley, A., Oliver, S., Sutcliffe, K., Rees, R., Brunton, G. and Kavanagh, J. (2004). Integrating qualitative research with trials in systematic reviews. *BMJ*, *328*(7446), 1010–1012. https://doi.org/10.1136/bmj.328.7446.1010.

Thomas, J., O'Mara-Eves, A. and Brunton, G. (2014). Using Qualitative Comparative Analysis (QCA) in systematic reviews of complex interventions: a worked example. *Systematic Reviews*, *3*(67).

Thomas, J., McDonald, S., Noel-Storr, A., Shemilt, I., Elliott, J., Mavergames, C. and Marshall, I. J. (2021). Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. *Journal of Clinical Epidemiology*, *133*, 140–151. https://doi.org/10.1016/j.jclinepi.2020.11.003.

Thorne, S. (2017). Metasynthetic madness: What kind of monster have we created? *Qualitative Health Research*, *27*(1), 3–12. https://doi.org/10.1177/1049732316679370.

Tricco, A. C., Antony, J., Zarin, W., Strifler, L., Ghassemi, M., Ivory, J., Perrier, L., Hutton, B., Moher, D. and Straus, S. E. (2015). A scoping review of rapid review methods. *BMC Medicine*, *13*(1). https://doi.org/10.1186/s12916-015-0465-6.

Tsafnat, G., Dunn, A., Glasziou, P., and Coiera, E. (2013). The automation of systematic reviews. *British Medical Journal*, *346*, f139–f139. https://doi.org/10.1136/bmj.f139.

van Grootel, L., van Wesel, F., O'Mara-Eves, A., Thomas, J., Hox, J. and Boeije, H. (2017). Using the realist perspective to link theory from qualitative evidence synthesis to quantitative studies: Broadening the matrix approach. *Research Synthesis Methods*, (March), 1–9. https://doi.org/10.1002/jrsm.1241.

Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, *39*(5), 426–431. https://doi.org/10.2307/3109916.

Young, K., Ashby, D., Boaz, A. and Grayson, L. (2002). Social Science and the Evidence-based Policy Movement. *Social Policy and Society*, *1*(3), 215–224. https://doi.org/10.1017/s1474746402003068.