

History of evidence synthesis to assess treatment effects: Personal reflections on something that is very much alive

Mike Clarke

Centre for Public Health, Institute of Clinical Sciences, Block B, Queens University Belfast, Royal Victoria Hospital, Grosvenor Road, Belfast BT12 6BA, Northern Ireland

Corresponding author: Mike Clarke, Email: m.clarke@qub.ac.uk

Setting the scene

It's a challenge to give an account of the 'history' of something that I have been a part of in recent decades. If I write about something that took place 25 years ago, is that historical, when it feels so near in time to me? When I teach undergraduates and show an article that was published before they were born, is that history or part of the here and now? And, when I look back on lectures from early in my career that mentioned things from the previous century, they felt, and were, some distance in the past. Writing this in August 2015, the last century is only 15 years behind us, and we need to refer to the century before last when we think about the 1800s. In less than 90 years, the 'present' will be the last 'century' and current and upcoming decades will be 'history'. That history is being made right now in relation to evidence synthesis. And the last couple of decades have seen developments that will become seen as pivotal.

In this essay, I try to capture some thoughts on the last hundred or more years, writing partly as a historian with a strong interest in how ideas evolve in parallel and independently, and also as someone who has been part of the history for 25 years and been fortunate to work with others who have been part of it for much longer. I will look at who and what happened 'early' rather than engage in a competition to find who was 'first' and try to provide a framework to help readers to think about history when it is still being made around us. I will try to highlight how several key elements for the successful conduct and uptake of evidence synthesis to assess treatment effects came together over recent decades to produce the upsurge in this activity, and a step change about 20 years ago. I draw on examples from the James Lind Library (www.JamesLindLibrary.org) as well as other accounts of various aspects of the history

of evidence synthesis in health and social care,^{1–6} including the influences of women in this history.⁷ This account should not itself be considered to be a 'systematic review'. It is a collection of illustrative examples to describe the journey that evidence synthesis has taken over more than 100 years, and to highlight examples of how the quality of this research has changed over time.^{8–12} I am sure that examples have been missed, some of which may be particularly important, and I should welcome information on any such examples and suggestions for improvements.

What does it mean?

There are many terms used for evidence synthesis, just as there are many terms for 'evidence'. This article focuses to a large extent on systematic reviews, in which a question is formulated, eligible studies are identified and appraised, and the findings are combined (sometimes mathematically) to summarise the effects, and perhaps to draw conclusions about the implications for future practice and research. The emphasis will be on research into the effects of interventions in health and social care, but it is important to note that there is a growing body of systematic reviews of other key areas for decision-making.¹² These include diagnostic accuracy and prognosis, and the use of evidence from other types of investigation including qualitative research, animal studies and modelling. This essay might, therefore, be considered to be a history of research synthesis with a focus on systematic reviews, and the important role played by a particular type of review: the Cochrane Review. The history of more statistical aspects of meta-analyses is dealt with in a companion article in the James Lind Library.¹³

An illustration of how historical analyses have been transformed by the living history of modern

developments is the work involved in the review of documents from the past; 30 years ago, someone wanting to know if a particular term had been used in 19th century medical journals would need to go to library, take the journals from the shelves and work through them methodically. Now, we go online and run a search of the digitised archives in seconds. This makes it much easier for us to find today's terms in the 19th century medical literature, but we still need to apply critical reasoning to consider whether the terms mean the same. This can make document review easier if a term was invented for the specific purpose of our interest and has no pre-history. This is the case with the term 'meta-analysis' but not with 'systematic review'. However, early uses of the latter can provide insight into why we use it now, and how people did similar things in the past. To begin this journey, a search of the *British Medical Journal* digitised archive for the phrase 'systematic review' finds an article from 1867 discussing the recently published edited reports of St Bartholomew's and St George's hospitals, which notes

Daunted by the difficulty of any systematic review of these collections of monographs, we shall only take a flying run through the pages; warning our readers, that they will do well to indemnify themselves by procuring the volumes for systematic perusal.¹⁴

This neatly captures the fact that preparing a systematic review can be a daunting task, and the challenges of undertaking one have been well described in guides to their conduct through recent decades.^{15–19}

Understanding the purpose of evidence syntheses, to understand why people do them

There are many reasons for doing evidence synthesis. These include the need to minimise bias by bringing together all of the available evidence on a particular topic, so that the emphasis is on the totality of the evidence and not merely a sample of the studies, highlighted because of their results. There is also a need to reduce the effects of the play of chance, by increasing the statistical power through the incorporation of as much data on the topic as possible, which can also be achieved by bringing together all of the available evidence on a particular topic but also requires that the data from that evidence can be combined mathematically, in meta-analyses. The history of the latter is dealt with partly in the companion article by Keith O'Rourke.¹³ Some of the reasons for doing evidence synthesis overlap, but some are mutually exclusive.

Some have changed in emphasis over time. However, the following list helps to orientate any work that wishes to look at why people have done and continue to do them. The examples that follow highlight some of these reasons, and these reasons help to provide a basis for understanding why an evidence synthesis, rather than a single study or a haphazard collection of studies, became so important:

- To organise a collection of the evidence;
- To appraise the quality of the evidence;
- To minimise bias, including avoiding undue emphasis on individual studies;
- To compare and contrast similar studies;
- To combine their findings, if possible and appropriate, to increase statistical power;
- To improve access to the evidence;
- To identify cost-effective interventions;
- To design better studies in the future.

As a starting point for considering the scientific value of evidence synthesis, let's go back to the 1880s and a presidential address to the British Association for the Advancement of Science by Lord Rayleigh²⁰ in Montreal. He said:

If, as is sometimes supposed, science consisted in nothing but the laborious accumulation of facts, it would soon come to a standstill, crushed, as it were, under its own weight. The suggestion of a new idea, or the detection of a law, supersedes much that has previously been a burden on the memory, and by introducing order and coherence facilitates the retention of the remainder in an available form. Two processes are thus at work side by side, the reception of new material and the digestion and assimilation of the old. One remark, however, should be made. The work which deserves, but I am afraid does not always receive, the most credit is that in which discovery and explanation go hand in hand, in which not only are new facts presented, but their relation to old ones is pointed out.

If we move forward 13 years, to another address to another meeting in North America, George Gould²¹ presented a vision to the first meeting of the Association of Medical Librarians in Philadelphia on 2 May 1898:

I look forward to such an organisation of the literary records of medicine that a puzzled worker in any part of the civilized world shall in an hour be able to gain a knowledge pertaining to a subject of the experience of every other man in the world.

These two examples from the late 19th century show both the scientific justification for evidence synthesis (to make best use of what has gone before) and the practical justification (to make it easier for decision-makers to access the knowledge from what has gone before). The latter also provides an important opportunity to note the important contribution that librarians and information specialists have made to improving access to the raw material for systematic reviews. In a review in the mid-1960s, Wechsler et al.²² report that they did ‘an extensive search of the literature’ for research evaluating antidepressant medications on hospitalised mental patients. Such searches have become easier over the subsequent half century, through the development of bibliographic databases containing millions of records and online access to full-text articles. In the early 1990s, when the Cochrane Collaboration was established (see below), the principal medical database, MEDLINE, contained fewer than 20,000 records that could be easily retrieved as reports of randomised trials.²³ Through an extensive programme of searching by members of Cochrane, and improved indexing, this number is now into the hundreds of thousands in MEDLINE and the Cochrane Central Register of Controlled Trials in the Cochrane Library, contains records for more than 880,000 reports.^{4,24}

James Lind: An early trial and early evidence synthesis

In his 1753 treatise on scurvy, not only did James Lind²⁵ describe his celebrated trial on scurvy but he also provided what the cover subtitle describes as a ‘Critical and Chronological View of what has been published on the subject’. He outlines the need for this with the words:

As it is no easy matter to root out prejudices...it became requisite to exhibit a full and impartial view of what had hitherto been published on the scurvy, and that in a chronological order, by which the sources of these mistakes may be detected. Indeed, before the subject could be set in a clear and proper light, it was necessary to remove a great deal of rubbish.

Other examples of efforts by researchers to summarise all the existing evidence are available in the James Lind Library from the decades at the start of the recent surge in activity. For example, in 1969, Smith et al.²⁶ wrote that their ‘comprehensive overview of antidepressant literature published in English...attempts to describe a total field of

research enquiry’. The Lind quote above captures one of the reasons for a key component of a modern evidence synthesis: the critical appraisal of the potentially eligible studies, with a view to minimising bias and separating the good from the bad. However, as noted by Chalmers et al., ‘It was not really until the 20th century...that the science of research synthesis as we know it today began to emerge’.² And, perhaps, it was not until nearly the fourth quarter of that century that proper recognition of evidence synthesis as ‘science’ began to develop, even though it has continued to be a challenge to have such research accepted as a scientific endeavour.

By way of illustration from the 1970s, in 1971, Feldman²⁷ wrote that systematically reviewing and integrating research evidence ‘may be considered a type of research in its own right – one using a characteristic set of research techniques and methods’. In the same year, Light and Smith²⁸ noted that it was impossible to address some hypotheses other than through analysis of variations among related studies, and that valid information and insights could not be expected to result from this process if it depended on the usual, scientifically undisciplined approach to reviews. Eugene Garfield²⁹ drew attention to the importance of scientific review articles in advancing original research, showing how review articles had high citation rates and review journals had high impact factors. He proposed a new profession, ‘scientific reviewer’, and his Institute for Scientific Information went on to co-sponsor (with Annual Reviews Inc.) an annual award for ‘Excellence in Scientific Reviewing’, administered by the National Academy of Sciences.³⁰

Mathematics, statistics and meta-analyses

One of the early examples cited by Chalmers et al.² of an evidence synthesis highlight how the use of statistical techniques helped to introduce scientific rigour to evidence synthesis. In the *British Medical Journal* of 5 November 1904, Karl Pearson, director of the Biometric Laboratory at University College London, pooled data from five studies of immunity and six studies of mortality among soldiers serving in India and South Africa to investigate the effects of a vaccine against typhoid. He calculated mean values across the two groups of study, noting:

Many of the groups in the South African experience are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved. Accordingly, it was needful to group them into larger series. Even thus the material appears to be so heterogeneous, and the results so

irregular, that it must be doubtful how much weight be attributed to the different results.³¹

In 1940, a group of researchers from Duke University in the U.S. produced the book *Extra-sensory perception after 60 years* which included statistical analyses that combined the results of individual studies and stated:

The comparison of the statistics of more than one experiment suggests a counterpart: the combination of them for an estimate of total significance.³²

It was in April 1976, though, that a key step took place with the introduction of a new term for this statistical combination: 'meta-analysis'. Gene Glass used his American Educational Research Association presidential address, to describe the need for better synthesis of the results of research studies, through a process he termed 'meta-analysis'. In the published version of the speech, he wrote:

My major interest currently is in what we have come to call – not for want of a less pretentious name – the meta-analysis of research. The term is a bit grand, but it is precise, and apt, and in the spirit of 'metamathematics', 'meta-psychology', and 'meta-evaluation'. Meta-analysis refers to the analysis of analyses.³³

Smith and Glass³⁴ published a substantial example of one such meta-analysis the following year, to look at research in psychotherapy. Their report drew on the accumulated evidence from 25,000 people in 375 studies of psychotherapy and counselling with 833 effect-size measures and was introduced with the words:

The purpose of the present research has three parts: (1) to identify and collect all studies that tested the effects of counseling and psychotherapy; (2) to determine the magnitude of effect of the therapy in each study; and (3) to compare the effects of different types of therapy and relate the size of effect to the characteristics of the therapy (e.g., diagnosis of patient, training of therapist) and of the study. Meta-analysis, the integration of research through statistical analysis of the analyses of individual studies,³³ was used to investigate the problem.³⁴

The term meta-analysis appears sporadically in the medical literature over the subsequent years but a notable example is in a 1982 comparison of 37 reports comparing pharmacological versus

non-pharmacological treatments for hypertension. Andrews et al.³⁵ wrote:

Glass introduced an approach called meta-analysis in which the properties of several studies could be recorded in quantitative terms and descriptive statistics applied to derive an overall conclusion. Thus, reviewing the published works ceases to require the judgment of Solomon and becomes a quasiempirical procedure. We used the meta-analytic technique to review non-pharmacological treatments for hypertension.

Around the same time as the introduction of the term 'meta-analysis', others were describing methods for combining the results of separate studies. In early 1977, Peto et al.³⁶ published the second in a pair of papers on the analyses of trials with prolonged follow up and the use of time-to-event analyses, showing how the results of separate studies might be combined as though each trial was a separate strata in a single study.

One of the things that subsequently accompanied these statistical techniques was a new way to display the findings of the meta-analyses: a graph that is now sometimes called the forest plot.¹ This shows the results for each study as a single line of data and graphical image, with a symbol at the bottom to indicate the overall average. Freiman et al.³⁷ displayed the results of 71 'negative' trials with horizontal lines for the confidence interval for each study and a mark to show the point estimate.

Lewis³⁸ produced something similar to display a meta-analysis of the effects of beta blockers on mortality. The Antiplatelet Trialists' Collaboration³⁹ published what would now be widely recognised as a forest plot in a systematic review of the prevention of vascular disease by antiplatelet therapy. This used squares of different sizes to show the weight of each study in the meta-analysis and the point estimates for the odds ratio from each trial, with the associated confidence intervals running through these. A rhombus, whose width was its confidence interval, provided the average at the bottom of the plot.³⁹

Systematic reviews as we know them today

In the month before Glass used the term 'meta-analysis' at the American Educational Research Association meeting, Shaikh et al.⁴⁰ published their article called a 'A systematic review of the literature on evaluative studies on tonsillectomy and adenoidectomy'. They outline their purpose as being:

to review the English language literature pertaining to evaluation of [tonsillectomy and adenoidectomy]

with a particular emphasis on an assessment of the scientific merit of studies which have attempted to determine the efficacy of this procedure.

A total of 28 reports describing 29 studies of tonsillectomy and adenoidectomy published between 1922 and 1970 were appraised and analysed, and the assessments of each study were presented in a table. This work reflects James Lind's intentions to separate the good from the bad, and to identify or overcome bias. This objective is distinct from the use of mathematical techniques to increase statistical power and decrease the effects of chance. Thus, Shaikh et al. provide a table showing how studies done by ear, nose and throat specialists were much more favourable to tonsillectomy and adenoidectomy (12 in favour, 0 against) than those done by public health or paediatric specialists (9 in favour, 8 against). In common with many of the challenges of the 2010s, Shaikh et al. conclude their review by calling for a well-conducted randomised trial to resolve the uncertainty⁴¹ and highlight how evidence of effectiveness is a key element in managing healthcare costs:⁴²

Aside from the high cost and lack of clear cut evidence of therapeutic efficacy, there is morbidity and mortality associated with tonsillectomy and adenoidectomy. . . . In view of the cost, financial and human, as well as the lack of evidence clearly supporting the continued performance of this procedure, it is suggested that a prospective, properly randomized controlled study be undertaken and that the methodologic pitfalls annotated in our review be guarded against. . . . In this era of escalating health care costs, society can only afford therapies which have been demonstrated to be of benefit.⁴⁰

This type of conclusion also serves to highlight the importance of doing reviews to provide the ethical, scientific, and economic and environmental justification when considering doing additional trials.⁴³⁻⁴⁵ This point was illustrated by Rogers and Clay,⁴⁶ who wrote that the results of their review of the existing trials 'suggest that the benefit of this drug in patients with endogenous depression who have not become institutionalized is indisputable, and that further drug-placebo trials in this condition are not justified'. Similarly, Baum et al.⁴⁷ concluded that a no-treatment control group should no longer be used in trials of antibiotic prophylaxis in colon surgery. A couple of other notable examples of evidence synthesis from the 1970s, which also cast doubt on the effects of interventions which may have looked promising when emphasis was given to the results of single studies are the work of Jan Stjernswärd⁴⁸

and Thomas Chalmers.⁴⁹ Stjernswärd⁵⁰ pooled the five-year survival results for five trials of postoperative radiotherapy for breast cancer, and concluded:

The routine use of postoperative irradiation in early breast cancer must be seriously questioned. Survival data argue against its use, despite the local effect on recurrence rates. If the routine use of prophylactic local radiotherapy after radical mastectomy were stopped, survival might increase and resources might be saved.

Chalmers⁵¹ brought together 14 trials of ascorbic acid for the common cold and combined the results from eight of them:

These are minor and insignificant differences, but in most studies the severity of symptoms was significantly worse in the patients who received the placebo. . . . All differences in severity and duration were eliminated by analyzing only the data from those who did not know which drug they were taking. Since there are no data on the long-term toxicity of ascorbic acid when given in doses of 1g or more per day, it is concluded that the minor benefits of questionable validity are not worth the potential risk, no matter how small that might be.

Collaboration and the 1980s

The following example from the start of the 1970s introduces the concept of the collaborative overview, in which researchers share their data. This need for researchers to collaborate together to ensure progress and reduce waste⁴⁴ had been highlighted in the 1950s by Kety.⁵² This approach to research synthesis became more common during the following decade. In 1970, in an early example of an individual participant data meta-analysis,⁵³ the International Anticoagulant Review Group⁵⁴ collected centrally and analysed original records for nearly 2500 patients from 9 of 10 identified trials to assess the effects of anticoagulant therapies after myocardial infarction. They wrote:

Although we recognised that the best solution would be a new collaborative controlled trial in a large number of patients, we decided that this was, at that time, quite impracticable. As a potentially useful and simple alternative we agreed on a systematic review of the data on individual patients pooled from all the adequately controlled trials that had been published recently.

Such collaborative efforts became a feature of some large scale reviews in the 1980s, in particular in other

areas of cardiovascular medicine and cancer.^{55,56} For instance, in October 1984, people responsible for randomised trials of tamoxifen or chemotherapy for the treatment of women with breast cancer met at Heathrow Airport in London to share findings and conduct a meta-analysis of the aggregate results.⁵⁷ They became the founders of the Early Breast Cancer Trialists' Collaborative Group (EBCTCG).^{58,59} In a short report in *The Lancet*, it was noted that:

Since the future treatment of many women might be importantly affected by this – or a further – overview of all available trials those meeting agreed to explore the possibility of extending their collaboration to include the central review of individual patient data.⁵⁷

Since then, the EBCTCG has conducted periodic reviews of the accumulating data from randomised trials of many aspects of the treatment of women with operable breast cancer, bringing further follow-up and additional trials into each cycle.^{60,61} The EBCTCG was recently used as an example of the successful sharing of participant-level data from clinical trials.⁶²

The spirit of collaboration to resolve uncertainties in healthcare in the 1980s extended beyond the establishment of groups of researchers willing to share individual participant data for collaborative meta-analyses. A notable example is the considerable international collaboration that led to the preparation of a large collection of systematic reviews of controlled trials relevant to perinatal care,^{63,64} and the use of electronic media to update and correct the reviews when necessary.⁶⁵ Looking back two decades later, Daniel Fox⁵ wrote:

The influence... on policy was mainly a result of... powerful blending of the rhetoric of scientific and polemical discourse, especially but not exclusively in ECPC; a growing constituency for systematic reviews as a source of 'evidence-based' health care among clinicians, journalists, and consumers in many countries; and recognition by significant policymakers who allocate resources to and within the health sector that systematic reviews could contribute to making health care more effective and to containing the growth of costs.

Cochrane Collaboration

Towards the end of the 1970s, in what might be considered to be a rallying call for evidence synthesis,⁶⁶ Archie Cochrane had written:

It is surely a great criticism of our profession that we have not organised a critical summary, by speciality

or subspeciality, adapted periodically, of all relevant randomised controlled trials.⁶⁷

At the end of the following decade, he used the phrase 'systematic review' in the foreword to the afore-mentioned compilation of evidence syntheses of maternity care interventions:

The systematic review of the randomised trials of obstetric practice that is presented in this book is a new achievement. It represents a real milestone in the history of randomised trials and in the evaluation of care, and I hope that it will be widely copied by other medical specialties.⁶⁸

Four years later, the international Cochrane Collaboration was established, following the opening of the first Cochrane Centre in Oxford, UK, in 1992.⁶⁹ The Cochrane Collaboration set itself the aim of helping people make well-informed decisions about healthcare by preparing, maintaining and promoting the accessibility of systematic reviews of the effects of healthcare interventions. It established an international infrastructure to support the production of systematic reviews across all areas of healthcare, with networks of individuals working together to prepare these reviews and keep them up to date. The advent of electronic publishing, which, at that time, meant publishing the material on floppy disks or compact disc read-only memory, allowed the full collection of systematic reviews to be provided to users on a regular basis, with the addition of new reviews and the updating of existing ones to take account of new evidence.

In 1995, the Collaboration's publishing partner, Update Software released the first issue of the *Cochrane Database of Systematic Reviews*.³ From 50 full *Cochrane reviews* in that first year, the number has grown to more than 6000 in 2015. The history of evidence synthesis took another major step in 1998, when the Database went onto the internet and, now, in its partnership with Wiley-Blackwell, the Collaboration publishes the full collection of reviews in the Cochrane Library online, with new and updated reviews appearing every few hours, rather than in quarterly or monthly bundles (www.cochranelibrary.com). The Collaboration itself has also grown considerably, from 77 people at the first Cochrane Colloquium in October 1993 to more than 30,000 in more than 100 countries (www.cochrane.org).⁷⁰

Growth

Although the Cochrane Collaboration remains the world's largest single producer of systematic reviews,

its output now accounts for only a small minority of the global output of evidence syntheses. Moher et al.¹² estimated that Cochrane reviews made up approximately 500 of the 2500 systematic reviews published each year. More recently, Bastian et al.⁴ used a variety of search strategies to show how steady growth in the number of evidence syntheses from the 1990s had transformed into a surge in recent years. Their graph clearly shows this, and it is important to note that what, at first sight, might look like a cumulative count of the number of systematic reviews found by the different types of search is actually the count for articles published in each single year, showing that, for non-Cochrane reviews in particular, each year saw more publications than the previous year. They estimated that 4000 reviews were being published annually by 2010 and predicted that this would continue to grow. This has been the case, and a search of PubMed in April 2015 finds 6313 articles published in 2014, using the Publication Type term *meta-analysis*. There are many more to come. For example, the international, prospective register of systematic reviews, PROSPERO, established in 2011⁷¹ is likely to surpass 10,000 records by the end of 2015 (www.crd.york.ac.uk/PROSPERO).

When the present and future have become history

I conclude by thinking forward to the next century. How will evidence synthesis in our current decades be viewed? What will be regarded as pivotal moments, step changes or gradual evolution? Some candidates that historians of the future might look to are:

- the increased use of prospective registries of trials to make it easier to find what trials have been done;⁷²
- increased automation of the systematic review process;⁷³
- greater access to the data from clinical trials⁶² and its use in individual participant data meta-analyses;⁷⁴
- greater use of material submitted to drug regulators;⁷⁵
- the use of new statistical techniques such as network meta-analyses;^{76,77}
- use of meta-epidemiology to improve the design and conduct of new studies;⁷⁸
- use of systematic reviews of animal research to inform research in humans;^{79,80}
- improvements in ways to summarise reviews and make them more accessible;^{81–83}
- the use of core outcome sets;⁸⁴
- the conduct of empirical research into the methods for doing research and reviews of these studies;⁸⁵ and
- perhaps, most importantly, even more recognition of the need for and benefits of systematic reviews as a way to justify and interpret new trials, and reduce waste.⁴⁴

The past hundred or more years have seen several developments in the science and practice of evidence synthesis. The last 20–30 years have seen important step changes in the numbers of these syntheses, and in the techniques to prepare and maintain them. The underpinning scientific rationale continues to resonate with the words of Lord Rayleigh.²⁰ The practical benefits of making it easier for people to make well-informed decisions and choices mean that Gould's²¹ vision of much improved access to knowledge and Kety's⁵² hope for greater collaboration among researchers may have been achieved.

Declarations

Competing interests: None declared

Funding: None declared

Ethical approval: Not applicable

Guarantor: MC

Contributorship: Sole authorship

Acknowledgements: I am grateful to the many people I have worked alongside who helped to create the recent history of evidence synthesis or are shaping its future. I am especially grateful to Iain Chalmers for his encouragement to write this article, and his improvements to the initial draft.

Provenance: Invited article from the James Lind Library.

References

1. Lewis S and Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001; 322: 1479–1480.
2. Chalmers I, Hedges LV and Cooper H. A brief history of research synthesis. *Eval Health Prof* 2002; 25: 12–37.
3. Starr M, Chalmers I, Clarke M and Oxman AD. The origins, evolution, and future of The Cochrane Database of Systematic Reviews. *Int J Tech Assess Health Care* 2009; 25(Suppl. 1): 182–195.
4. Bastian H, Glasziou P and Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010; 7: e1000326.
5. Fox DM. Systematic reviews and health policy: the influence of a project on perinatal care since 1988. *Milbank Q* 2011; 89: 425–449.
6. Shadish WR and Lecy JD. The meta-analytic big bang. *Res Synth Methods* 2015; 6: 246–264.
7. Dickersin K. Innovation and cross-fertilization in systematic reviews and meta-analysis: the influence of women investigators. *Res Synth Methods* 2015; 6: 277–283.

8. Kass EH. Reviewing reviews. In: Warren KS (ed.) *Coping with the Biomedical Literature*. New York: Praeger, 1981, pp.79–91.
9. Mulrow CD. The medical review article: state of the science. *Ann Int Med* 1987; 106: 485–488.
10. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA and Chalmers TC. Meta-analysis of randomized controlled trials. *New Engl J Med* 1987; 316: 450–455.
11. Mulrow CD. We've come a long way, baby! *The Cochrane Collaboration Methods Working Groups Newsletter*, 1–2, 1998. See http://community.cochrane.org/sites/default/files/uploads/Newsletters/MGNews_1998.pdf.
12. Moher D, Tetzlaff J, Tricco AC, Sampson M and Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; 4: e78.
13. O'Rourke K. A historical perspective on meta-analysis: dealing quantitatively with varying study results. *JLL Bull*, 2006. See <http://www.jameslindlibrary.org/articles/a-historical-perspective-on-meta-analysis-dealing-quantitatively-with-varying-study-results/>.
14. Anon. Reviews and notices. *BMJ* 1867; 2: 425–426.
15. Goldschmidt PG. Information synthesis: a practical guide. *HSR* 1986; 21: 215–236.
16. Cooper H, Hedges LV, ed. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, 1994.
17. Pettiti DB. *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*. New York: Oxford University Press, 1994.
18. Higgins JPT and Green S, ed. *Cochrane Handbook for Systematic Reviews of Interventions*. Oxford: The Cochrane Collaboration; John Wiley and Sons Ltd., 2008.
19. Centre for Reviews and Dissemination (CRD). *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. York: University of York, 2009.
20. Rayleigh L. Address by the Rt. Hon. Lord Rayleigh. In: *Report of the fifty-fourth meeting of the British Association for the Advancement of Science*, Montreal, QC, Canada, August and September. London: John Murray, 1885.
21. Gould GM. The work of an association of medical librarians. *J Med Libr Assoc* 1898; 1: 15–19.
22. Wechsler H, Grosser GH and Greenblatt M. Research evaluating antidepressant medications on hospitalized mental patients: a survey of published reports during a five-year period. *J Nerv Ment Dis* 1965; 141: 231–239.
23. Dickersin K, Scherer R and Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286–1291.
24. Lefebvre C, Eisinga A, McDonald S and Paul N. Enhancing access to reports of randomized trials published world-wide – the contribution of EMBASE records to the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library. *Emerg Them Epidemiol* 2008; 5: 13.
25. Lind J. *A Treatise of the Scurvy. In Three Parts. Containing an Inquiry into the Nature, Causes and Cure, of that Disease. Together with a Critical and Chronological View of What has been Published on the Subject*. Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson, 1753.
26. Smith A, Traganza E and Harrison G. Studies on the effectiveness of antidepressant drugs. *Psychopharmacol Bull* 1969; March(Suppl.): 1–53.
27. Feldman KA. Using the work of others: some observations on reviewing and integrating. *Sociol Educ* 1971; 44: 86–102.
28. Light RJ and Smith PV. Accumulating evidence: procedures for resolving contradictions among research studies. *Harvard Educ Rev* 1971; 41: 429–471.
29. Garfield E. Proposal for a new profession: scientific reviewer. *Scientist* 1977; 3: 84–87.
30. Garfield E. The NAS James Murray Luck Award for excellence in scientific reviewing. *Scientist* 1979; 4: 127–131.
31. Pearson K. Report on certain enteric fever inoculation statistics. *BMJ* 1904; 3: 1243–1246.
32. Pratt JG, Rhine JB, Smith BM, Stuart CE and Greenwood JA. *Extra-Sensory Perception after Sixty Years: A Critical Appraisal of the Research in Extra-Sensory Perception*. New York: Henry Holt, 1940.
33. Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976; 10: 3–8.
34. Smith ML and Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977; 32: 752–760.
35. Andrews G, MacMahon SW, Austin A and Byrne DG. Hypertension: comparison of drug and non-drug treatments. *BMJ* 1982; 284: 1523–1526.
36. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer* 1977; 35: 1–39.
37. Freiman JA, Chalmers TC, Smith H and Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials. *New Engl J Med* 1978; 299: 690–694.
38. Lewis JA. Beta-blockade after myocardial infarction – a statistical view. *Br J Clin Pharmacol* 1982; 14: 15S–21S.
39. Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged anti-platelet treatment. *BMJ* 1988; 296: 320–331.
40. Shaikh W, Vayda E and Feldman W. A systematic review of the literature on evaluative studies on tonsillectomy and adenoidectomy. *Pediatrics* 1976; 57: 401–407.
41. Clarke L, Clarke M and Clarke T. How useful are Cochrane reviews in identifying research needs? *J Health Serv Res Pol* 2007; 12: 101–103.
42. Garner S, Docherty M, Somner J, Sharma T, Choudhury M, Clarke M, et al. Reducing ineffective practice: challenges in identifying low-value health care using Cochrane systematic reviews. *J Health Serv Res Pol* 2013; 18: 6–12.

43. Clarke M. Doing new research? Don't forget the old: nobody should do a trial without reviewing what is known. *PLoS Med* 2004; 1: 100–102.
44. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. *Lancet* 2014; 383: 101–104.
45. Clarke M, Brice A and Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS ONE* 2014; 9: e102670.
46. Rogers SC and Clay PM. A statistical review of controlled trials of imipramine and placebo in the treatment of depressive illnesses. *Br J Psychiatry* 1975; 127: 599–603.
47. Baum ML, Anish DS, Chalmers TC, Sacks HS, Smith H and Fagerstrom RM. A survey of clinical trials of antibiotic prophylaxis in colon surgery: evidence against further use of no-treatment controls. *New Engl J Med* 1981; 305: 795–799.
48. Stjernswärd J. Personal reflections on contributions to pain relief, palliative care and global cancer control. *JLL Bull.* See <http://www.jameslindlibrary.org/articles/personal-reflections-on-contributions-to-pain-relief-palliative-care-and-global-cancer-control/> (2013, last checked 3 March 2016).
49. Dickersin K and Chalmers F. Thomas C Chalmers (1917–1995): a pioneer of randomized clinical trials and systematic reviews. *JLL Bull.* See <http://www.jameslindlibrary.org/articles/thomas-c-chalmers-1917-1995/> (2014, last checked 2 March 2016).
50. Stjernswärd J. Decreased survival related to irradiation postoperatively in early breast cancer. *Lancet* 1974; 304: 1285–1286.
51. Chalmers TC. Effects of ascorbic acid on the common cold. An evaluation of the evidence. *Am J Med* 1975; 58: 532–536.
52. Kety S. Comment. In: Cole JO and Gerard RW (eds) *Psychopharmacology. Problems in Evaluation*. Washington, DC: National Academy of Sciences, Publication 583, 1959, pp.651–652.
53. Clarke M and Godwin J. Systematic reviews using individual patient data: a map for the minefields? *Ann Oncol* 1998; 9: 827–833.
54. International Anticoagulant Review Group. Collaborative analysis of long-term anti-coagulant administration after acute myocardial infarction. *Lancet* 1970; 1: 203–209.
55. Stewart L and Clarke M for the Cochrane Collaboration Working Group on meta-analyses Using Individual Patient Data. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Stat Med* 1995; 14: 2057–2079.
56. Clarke M, Stewart L, Pignon J-P and Bijlens L. Individual patient data meta-analyses in cancer. *Br J Cancer* 1998; 77: 2036–2044.
57. Anon. Review of mortality results in randomized trials in early breast cancer. *Lancet* 1984; 2: 1205.
58. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). *Treatment of Early Breast Cancer. Vol 1. Worldwide Evidence, 1985–1990*. Oxford: Oxford University Press, 1990.
59. Darby S, Davies C and McGale P. The Early Breast Cancer Trialists' Collaborative Group: a brief history of results to date. In: Davison AC, Dodge Y and Wermuth N (eds) *Celebrating Statistics*. Oxford: Oxford University Press, 2005, pp. 185–198.
60. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. an overview of 61 randomized trials among 28,896 women. *New Engl J Med* 1988; 319: 1681–1692.
61. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* 2012; 379: 432–444.
62. Varnai P, Rentel MC, Simmonds P, Sharp TA, Mostert B and de Jongh T. Assessing the research potential of access to clinical trial data. A report to the Wellcome Trust. Brighton: technopolis group. www.wellcome.ac.uk/stellent/groups/corporatesite/msh_peda/documents/web_document/WTP058912.pdf (2015, last checked 17 Mar 2017).
63. Chalmers I, Enkin M and Keirse MJNC. *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press, 1989.
64. Sinclair JC and Bracken MB. *Effective Care of the Newborn Infant*. Oxford: Oxford University Press, 1992.
65. Chalmers I, ed. *The Oxford Database of Perinatal Trials*. Oxford: Oxford University Press, 1988.
66. Chalmers I. Archie Cochrane (1909–1988). *JLL Bull.* See <http://www.jameslindlibrary.org/articles/archie-cochrane-1909-1988/> (2016, last checked 3 March 2016).
67. Cochrane AL. 1931–1971: a critical review, with particular reference to the medical profession. In: Teeling-Smith G (ed.) *Medicines for the Year 2000*. London: Office of Health Economics, 1979, pp.1–11.
68. Cochrane AL. Foreword. In: Chalmers I, Enkin M and Keirse MJNC (eds) *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press, 1989.
69. Chalmers I, Dickersin K and Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMJ* 1992; 305: 786–788.
70. Allen C and Richmond K. The Cochrane collaboration: international activity within Cochrane review groups in the first decade of the twenty-first century. *J Evid Base Med* 2011; 4: 2–7.
71. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. PROSPERO at one year: an evaluation of its utility. *Syst Rev* 2013; 2: 4.
72. Ghersi D and Pang T. From Mexico to Mali: four years in the history of clinical trial registration. *J Evid Base Med* 2009; 2: 1–7.
73. Adams CE, Polzmacher S and Wolff A. Systematic reviews: work that needs to be done and not to be done. *J Evid Base Med* 2013; 6: 232–235.

74. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015; 313: 1657–1665.
75. Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. *Cochrane Database Syst Rev* 2014; 4: CD008965.
76. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol* 2014; 67: 138–143.
77. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Int Med* 2015; 162: 777–784.
78. Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized controlled trials. *Ann Int Med* 2012; 157: 429–438.
79. Sena ES, Briscoe CL, Howells DW, Donnan GA, Sandercock PA and Macleod MR. Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion models of stroke: systematic review and meta-analysis. *J Cereb Blood Flow Metab* 2010; 30: 1905–1913.
80. Sena ES, Currie GL, McCann SK, Macleod MR and Howells DW. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. *J Cereb Blood Flow Metab* 2014; 34: 737–742.
81. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011; 64: 383–394.
82. Murthy L, Shepperd S, Clarke MJ, Garner SE, Lavis JN, Perrier L, et al. Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians. *Cochrane Database Syst Rev* 2012; 9: CD009401.
83. Maguire LK and Clarke M. How much do you need: a randomised experiment of whether readers can understand the key messages from summaries of Cochrane Reviews without reading the full review. *J R Soc Med* 2014; 107: 444–449.
84. Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS ONE* 2014; 9: e99111.
85. Anon. Education section – studies within a review (SWAR). *J Evid Base Med* 2012; 5: 188–189.