

ANALYSIS



The knowledge system underpinning healthcare is not fit for purpose and must change

The medical literature is biased and inundated with poor quality trials. **Ian Roberts and colleagues** explain how these problems affect systematic reviews and how they might be overcome

Ian Roberts *professor of epidemiology and public health*, Katharine Ker *lecturer in epidemiology*, Phil Edwards *statistician*, Deirdre Beecher *information scientist*, Daniela Manno *clinical lecturer*, Emma Sydenham *managing editor*

Cochrane Injuries Group, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Information on the effectiveness and safety of healthcare should be valid, precise, up to date, clear, and freely available. Currently none of these criteria are fully satisfied, and Cochrane systematic reviews are not the solution. In this article we explain why the knowledge system for healthcare is unfit for purpose and suggest how it should change.

The problem with systematic reviews

Because the medical literature contains a biased sample of trials, systematic reviews that are based on it are also biased.¹ Despite decades of exhortation about trial publication, about half of all trials are unpublished, and even the most diligent efforts to synthesise the results from all (or an unbiased sample of) relevant trials are in vain.^{1,2} Even when trials are identified selective outcome reporting limits their validity.^{3,4} The literature is awash with low quality, underpowered, single centre trials and the trend is upwards. As Altman recognised over 20 years ago “much poor research arises because researchers feel compelled for career reasons to carry out research that they are ill equipped to perform, and nobody stops them.”⁵

The systematic review movement is renowned for its obsessive zeal to find every published trial, irrespective of size or quality. These efforts have increased the accessibility of many trials that should never have been conducted. Their citation in reviews perpetuates the problem by contributing to the impact factor of the journals that published them. In the United Kingdom, the funding provided to Cochrane review groups is proportional to the number of trials included in reviews, creating a financial incentive to find and include every trial regardless of its quality.

Most journal editors and systematic reviewers take trial reports at face value with little or no effort to confirm whether a particular trial even took place. A Cochrane review showing that high dose mannitol reduced the risk of death after head injury was retracted after the review group editors were unable to confirm that any of the included trials took place.⁶ The

conclusions of a systematic review of starch solutions in critically ill patients changed substantially after excluding seven trials that were retracted owing to misconduct by an investigator.^{7,8} Investigating possible fraud is hard work, and it is easier for journal editors to ignore the problem and perpetuate the myth that peer review of trial reports ensures their scientific quality. As part of the investigation of the high dose mannitol trials, the Cochrane Injuries Group editors contacted the editor of the journal that published one of the doubtful trials. He responded, “As you can tell by Dr Marshall’s editorial, we all doubted the data. But to doubt is different from concluding that Dr Cruz fabricated the data. I thought he did, but hoped as stated in the editorial that publication would encourage repetition of the studies. My editorial board thought Dr Cruz’s work should be published. I wouldn’t trust the data.”⁶ How can Cochrane claim to provide trusted evidence when all evidence is taken on trust?

The median number of trials in Cochrane reviews is between six and 16, and the median number of patients per trial is about 80.⁹ Consequently, most meta-analyses include small numbers of trial participants. Meta-analyses with sparse data can miss modest but clinically important treatment effects, and the potential for random error to result in false positive conclusions is considerable.¹⁰⁻¹³ Repeated statistical testing as trials accumulate increases this potential. Statistical analyses that account for multiple testing show that as many as two thirds of apparently conclusive findings in Cochrane reviews could be falsely positive.¹⁴ Clinical trial protocols invariably include an estimate of the sample size needed for the results to be reliable. But systematic reviewers rarely estimate how many participants would need to be included in a meta-analysis for reliable results.¹⁵

Distilling therapeutic truth from a ferment of poor quality trials is a challenge. But as statisticians and methodologists have risen to it, conducting reviews has become complex and time

Correspondence to: I Roberts ian.roberts@lshtm.ac.uk

Extra material supplied by the author (see <http://www.bmj.com/content/350/bmj.h2463?tab=related#datasupp>)

consuming, and reviews have become increasingly unreadable. Many reviews are longer than the combined length of the included trial reports.⁹ With the resources available and the increased complexity of the task, it is impossible to keep up to date with new trial research, and most Cochrane reviews are many years out of date.⁹

What can be done?

Exclude unregistered trials

Prospective clinical trial registration was first proposed in 1986, although it took until 2005 for the International Committee of Medical Journal Editors to make registration a prerequisite for publication.^{16 17} A decade later, less than a third of journals insist on registration.¹⁸ Since January 2015 all new reviews and updates undertaken by the Cochrane Injuries Group—which is responsible for the publication of systematic reviews in the Cochrane Library related to the prevention, treatment, and rehabilitation of traumatic injury—now only include prospectively registered trials, unless the trial was published before 2010. This strategy is unlikely to exclude any adequately powered high quality trials. We urge other Cochrane groups and reviewers to do the same to avoid being complicit in raising the profile of substandard research. Reviews of non-randomised studies are also vulnerable to bias from selective reporting, but until there are more determined efforts to tackle this problem, through prospective registration of protocols, our proposals can only be applied to randomised trials.¹⁹

As a pilot test, we retrospectively applied our new policy to the Cochrane review on colloids versus crystalloids for fluid resuscitation in critically ill patients, which includes 78 randomised controlled trials and makes three comparisons.²⁰ The data supplement on thebmj.com shows the number of included trials and point estimates of risk ratio for all-cause mortality (measured at final follow-up in each trial) in the original paper (pre-policy), after exclusion of trials published after 1 January 2010 that were not prospectively registered (post-policy), and when all trials that were not prospectively registered were excluded. The risk ratios for the effects of each comparison were essentially unchanged, even when all trials that were not prospectively registered were excluded. We judged only one of the trials excluded to be of high quality.

Statistical checks

The risk of fraud cannot be ignored. Since January 2015 we have also carried out statistical checks on trials included in our systematic reviews. Authors of doubtful trials are asked to provide the original data for checking. If they decline the trial is removed from the analysis and the reasons are stated in the review.

Notably, the Cochrane review of starch solutions in critically ill patients that included several flawed trials had been flagged by Ioannidis and colleagues as showing extreme homogeneity between studies.^{20 21 22} They contacted the author of the suspect trials and were reassured that they were real. No further action was taken. Although we cannot be sure that our policy will substantially reduce the number of questionable trials included in reviews, it would be inappropriate not to act when there are reasonable grounds for suspicion.^{6 22}

Sample size estimate

The number of trial participants needed for a reliable meta-analysis should be at least as large as for an appropriately powered trial.^{13 15} The Cochrane Injuries Group requires authors

to specify an “information size” estimate (based on plausible treatment effects) as part of the protocol for a systematic review (box). If the number of participants included in the meta-analysis is lower than the estimated information size, the reliability of all primary outcomes with “statistically significant” treatment effects will be further explored using sequential analysis methods.

Eliminating dependence on publication

One of the most important contributions of The Cochrane Collaboration was to show the world that the knowledge base for healthcare decision making is inadequate. Thousands of articles have been published about publication bias. However, the challenge is not to describe the flaws in the current system but to create a better one, where decisions about healthcare are informed by valid and reliable evidence.

Clinical trial information is too important to depend on the publication game. The solution to publication bias is to eliminate the dependence on publication. But the clear stream of reason that motivated trial registration has been lost in the dreary desert sands of habitual database searching. When reviews include only prospectively registered trials, the burden of conducting reviews will be reduced. Because many words are devoted to describing poor quality trials and the efforts made to separate the methodological wheat from the chaff, the exclusion of chaff will make reviews more readable. We should look to a future when reviews are conducted and updated by selecting and downloading data from clinical trial registries.

Contributors and sources: The authors are members of the core editorial team of the Cochrane Injuries Group, which is responsible for the publication of systematic reviews in the Cochrane Library related to the prevention, treatment, and rehabilitation of traumatic injury, including emergency resuscitation. The ideas presented here were developed in response to important matters that have affected the group's reviews over the past five years, such as how to handle potentially fraudulent research. IR drafted the manuscript. All authors gave substantial contributions to the conception of the paper and revised it critically for important intellectual content. All authors approved the final version of the paper for publication. IR is guarantor. This article reflects the views of the authors and not Cochrane.

Competing interests: All authors have read and understood BMJ policy on declaration of interests and declare the following interests: all authors are members of the Cochrane Injuries Group Editorial Base, which is funded by the National Institute for Health Research. They received no support from any organisation for the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990;263:1405-8.
- 2 Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:1-193.
- 3 Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.
- 4 Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383:166-75.
- 5 Altman DG. The scandal of poor medical research. *BMJ* 1994;308:283-4.
- 6 Roberts I, Smith R, Evans S. Doubts over head injury studies. *BMJ* 2007;334:392-4.
- 7 Bunn F, Alderson P, Hawkins V. Colloid solutions for fluid resuscitation. *Cochrane Database Syst Rev* 2001;2:CD001319.
- 8 Zarychanski R, Abou-Setta AM, Turgeon AF, et al. Association of hydroxyethyl starch administration with mortality and acute kidney injury in critically ill patients requiring volume resuscitation: a systematic review and meta-analysis. *JAMA* 2013;309:678-88.
- 9 Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7:e1000326.
- 10 Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials* 1997;18:580-93.
- 11 Thorlund K, Imberger G, Walsh M, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis—a simulation study. *PLoS One* 2011;6:e25491.

Estimating the information size for a systematic review

The information size for a meta-analysis can be approximated by the sample size that would be needed for a single randomised controlled trial to detect the hypothesised intervention effect. For example, to detect a reduction in the risk of an event from 10% to 8.5% (that is, a risk ratio of 0.85) with 90% power at the 5% significance level, a randomised controlled trial would need to include about 16 000 participants (8 000 participants in each arm).

The pooled estimate of an intervention effect obtained from a meta-analysis of two or more studies is a weighted average of the individual study results. However, statistical heterogeneity between trials increases the standard error of the pooled estimate. For this reason, the information size for a meta-analysis is usually larger than the sample size of a single trial.

An information size estimate that takes into account the expected heterogeneity between trials can be derived using sequential analysis software.²³ Using the example above, statistical heterogeneity equivalent to an I-squared of 25% or 50% would increase the information size from 16 000 to 21 000 or 32 000, respectively.

Key messages

Because the medical literature is biased, systematic reviews based on it are also biased. Many reviews are out of date and unreadably long

Including only prospectively registered trials in systematic reviews will improve validity and readability

Insisting that authors of doubtful trials provide the original trial data for statistical checking will improve validity

Requiring review authors to specify an estimated information size based on plausible treatment effects will reduce the risk of false positive conclusions and make reviews more reliable

Trial registries should include full protocols and datasets to facilitate the conduct of valid systematic reviews

- 12 AlBalawi Z, McAlister FA, Thorlund K, Wong M, Wetterslev J. Random error in cardiovascular meta-analyses: how common are false positive and false negative results? *Int J Cardiol* 2013;168:1102-7.
- 13 Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008;61:763-9.
- 14 Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive—trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *Int J Epidemiol* 2009;38:287-98.
- 15 Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet* 1998;351:47-52.
- 16 Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529-41.
- 17 De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *Lancet* 2004;364:911-2.
- 18 Wager E, Williams P, Project Overcome Failure to Publish Negative Findings Consortium. "Hardly worth the effort"? Medical journals' policies and their editors' and publishers' views on trial registration and publication bias: quantitative and qualitative study. *BMJ* 2013;347:f5248.
- 19 Loder E, Groves T, MacAuley D. Registration of observational studies. *BMJ* 2010;340:c950.
- 20 Perel P, Roberts I, Ker K. Colloids versus crystalloids for fluid resuscitation in critically ill patients. *Cochrane Database Syst Rev* 2013;2:CD000567.
- 21 Ioannidis JP. Scientific inbreeding and same-team replication: type D personality as an example. *J Psychosom Res* 2012;73:408-10.
- 22 Ioannidis JP. Meta-analyses of hydroxyethyl starch for volume resuscitation. *JAMA* 2013;309:2209.
- 23 Thorlund K, Engström J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for Trial Sequential Analysis (TSA). 2011. www.ctu.dk/tsa/files/tsa_manual.pdf.

Accepted: 25 March 2015

Cite this as: *BMJ* 2015;350:h2463

© BMJ Publishing Group Ltd 2015