

Key concepts for informed health choices. 2.2: reviews of the effects of treatments should be fair

AD Oxman¹, I Chalmers² and A Dahlgren³

¹Centre for Epidemic Interventions Research, Norwegian Institute of Public Health, 0213 Oslo, Norway

²Centre for Evidence-Based Medicine, University of Oxford, OX2 6GG, UK

³Faculty of Health Sciences, Oslo Metropolitan University, 0130 Oslo, Norway

Corresponding author: AD Oxman. Email: oxman@online.no

This is the second of four essays in this series explaining key concepts about the trustworthiness of evidence from treatment comparisons. In this essay, we explain four considerations about reviews of the effects of treatments – considering whether:

- systematic methods were used,
- unpublished results were considered,
- treatments were compared across studies and
- important assumptions were tested.

The basis for these concepts is described elsewhere.¹

Consider whether systematic methods were used

A systematic review is a summary of research evidence (studies) that uses systematic and explicit methods to summarise the research on the effects of a treatment (or some other topic). A systematic review addresses a clearly formulated question using a structured approach to identify, select and critically appraise relevant studies, and to collect and analyse data from the studies that are included in the review. Systematic reviews begin with protocols, which should be registered and searchable in registries such as Prospero.²

Even reviews that purport to be systematic may not be. Reviews that do not use systematic methods may result in biased or imprecise estimates of the effects of treatments because the selection of studies for inclusion may be biased, or the methods may result in some studies not being found. In addition, the appraisal of the quality of some studies may be biased, or the synthesis of the results of the selected studies may be inadequate or inappropriate.

For example, if a systematic review of giving blood thinners to patients with an acute heart attack had been done in the late 1970s, it would have established the effectiveness of that treatment

about 10 years before the results of a very large randomised trial became available.³ If those results had been acted upon, thousands of premature deaths could have been avoided. Instead, recommendations were based on unsystematic reviews of the evidence. Similarly, the harmful effects of medicines to reduce heart rhythm abnormalities in patients with an acute heart attack could have been recognised years earlier. And thousands of deaths caused by those medicines could have been prevented if those results had been acted upon.

Consider whether unpublished results were considered

Many fair comparisons are never published, and outcomes are sometimes left out from those that are published. Those that are published are more likely to report favourable results. Consequently, reliance on published reports alone sometimes results in the beneficial effects of treatments being overestimated and the adverse effects being underestimated.

For example, among trials of antidepressant drugs submitted to the U.S. Federal Drug Administration (FDA) or the Swedish drug regulatory authority, efficacy trials reporting positive results and larger effect sizes were more likely to be published subsequently. A review of trials supporting new medicines approved by the FDA between 1998 and 2000 found that over half of all supporting trials for FDA-approved drugs remained unpublished for five or more years after approval.⁴ Selective reporting of trial results was found for commonly marketed medicines.

Biased under-reporting of research is a major problem that is far from being solved. It is scientific and ethical malpractice and wastes research resources. Selective reporting is an important reason why fair comparisons of treatments should begin with protocols that are registered and searchable in registries such as clinicaltrials.gov. This can also help to

reduce selective reporting of some outcomes but not others in published reports, depending on the nature and direction of the results.

Consider whether treatments were compared across studies

For many conditions (e.g. depression) there are more than two possible treatments (for example, different medicines, or types of psychotherapy). Only very rarely are all the possible treatments for a condition compared in a single study, so it may be necessary to consider indirect comparisons among treatments. For example, there may be comparisons of drug A with placebo and comparisons of drug B with placebo, but no studies that compare drug A with drug B directly. In this case, indirect comparisons among studies may be needed to inform a decision about whether to use drug A or drug B. However, there can be important differences between the studies examined in addition to the treatments they assessed, for example, differences in characteristics of the participants, or the way the comparisons were done, or in the outcome measures used. These differences can result in misleading estimates of treatment effects.

A systematic review of different doses of aspirin illustrates the problem with indirect comparisons.⁵ The authors found five randomised trials that compared aspirin with placebo to prevent graft occlusion after coronary artery bypass surgery. Two trials tested medium-dose and three low-dose aspirin. Based on the indirect comparison, the relative risk reduction for medium-dose compared to low-dose aspirin was 0.74 (95% confidence interval 0.52 to 1.06; $P=0.10$), suggesting the possibility of a larger effect with medium-dose aspirin. However, there are other characteristics of the trials that might be responsible for any differences found (or undetected differences that might exist). Compared with the low-dose trials, the patients included in the medium-dose trials may be different, interventions other than aspirin may have been differently administered and outcomes may have been measured differently (e.g. dissimilar criteria for occlusion or different durations of follow-up). Differences in study methods and the risk of bias may also explain the results.

Consider whether important assumptions were tested

Sometimes treatment claims are based on chains of evidence, or models. For example, the effects of using a diagnostic test may depend on how accurate the

test is, assumptions about what will be done based on the test results and evidence of the effects of what is done. Similarly, evidence of the effects of public health and health system policies sometimes comes from models that combine different types of studies and assumptions; and assumptions are sometimes made when fair comparisons are combined in systematic reviews. When treatment comparisons depend on assumptions, it is important to consider their basis and to test how sensitive the results are to plausible changes in the assumptions made. For example, a model used to compare the effects of using different diagnostic tests on outcomes that are important to patients might require assumptions about what actions doctors or patients will take, based on test results. If that is uncertain, it is important to consider whether changing the assumptions has a substantial impact on the estimated difference in outcomes important to patients.

During and prior to the COVID-19 pandemic, there have been few randomised trials of public health measures used to control spread of infections, such as school closures.⁶ As a result, estimates of the effects of those interventions have frequently been based on models and non-randomised studies. The modelling studies make many different assumptions and often suggest different effects. For example, some modelling studies have suggested that school closures can reduce community transmission of the coronavirus, while others disagree.⁷ These models depend on many assumptions, and changes in these assumptions can change the results. Different models make different assumptions about per-contact transmission probabilities, how many parents go to work or work at home when schools are closed or opened, changes in contacts outside of home because of schools closing or opening, what other protective measures are in place, what happens during holidays, what proportion of infected people have symptoms, how long they are infected before they have symptoms and are tested, how long the symptoms last, contact tracing, how many people without symptoms are tested, the accuracy of testing, delays in getting test results, and compliance with and effects of isolation and quarantine. Because of all these assumptions and important uncertainty about many of them, the results of these modelling studies are very uncertain.

Early in the pandemic, some assumptions were empirically informed, such as how populations are distributed spatially. However, other assumptions were seemingly anecdotal, such as an assumption that children were twice as likely as adults to transmit the coronavirus. That assumption helped justify school closures. However, subsequent epidemiological studies suggested, if anything, children may be

less likely to transmit the virus.⁸ In addition, some models did not consider health consequences beyond deaths from coronavirus or how social and economic consequences might affect health. Models can be helpful when there is extreme uncertainty, but it is important to recognise their limitations and uncertainty.

Implications

- Whenever possible, use up-to-date systematic reviews of fair comparisons to inform decisions rather than non-systematic reviews of fair comparisons of treatments.
- Be aware of the possibility of biased underreporting of fair comparisons and assess whether the authors of systematic reviews have addressed this risk.
- Indirect comparisons are sometimes needed to inform treatment choices. In these circumstances, careful consideration should be given to differences between the studies besides the treatments that were compared.
- Whenever treatment comparisons depend on assumptions, consider whether the assumptions are well-founded and how sensitive the results are to plausible changes in the assumptions that are made.

Declarations

Competing Interests: None declared.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Research Council of Norway (Project numbers 220603/H10 and 284683). The funder had no role in the decision to publish, or preparation of the manuscript.

Ethics approval: Not applicable.

Guarantor: ADO.

Contributorship: ADO, IC and AD conceptualised, reviewed and edited drafts of this essay. ADO prepared the first draft.

Provenance: Not commissioned; invited article from the James Lind Library.

Note: Additional material for this article is available from the James Lind Library website [www.jameslindlibrary.org], where it was previously published

References

1. Oxman AD, Chalmers I, Dahlgren A and the Informed Health Choices Group. Key concepts for informed health choices: a framework for enabling people to think critically about health claims (Version 2022). *IHC Working Paper*, 2022. <http://doi.org/10.5281/zenodo.6611932>
2. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst Rev* 2012; 1: 2.
3. Antman EM, Lau J, Kupelnick B, Mosteller F and Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992; 268: 240–248.
4. Lee K, Bacchetti P and Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. *PLoS Med* 2008; 5: e191.
5. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence – indirectness. *J Clin Epidemiol* 2011; 64: 1303–1310.
6. Glasziou PP, Michie S and Fretheim A. Public health measures for covid-19. *BMJ* 2021; 375: n2729.
7. Walsh S, Chowdhury A, Braithwaite V, Russell S, Birch JM, Ward JL, et al. Do school closures and school reopenings affect community transmission of COVID-19? A systematic review of observational studies. *BMJ Open* 2021; 11: e053371.
8. Reddy S. How epidemiological models fooled us into trusting bad assumptions. *Barrons*. 29 April 2020. See www.barrons.com/articles/the-danger-of-overreliance-on-epidemiological-models-51588179008 (last checked 9 November 2022).