The Royal Society of Medicine

Journal of the Royal Society of Medicine; 2022, Vol. 115(10) 408–411
DOI: 10.1177/01410768221135497

Key concepts for informed health choices. 1.2: Seemingly logical assumptions about research can be misleading

AD Oxman[1], I Chalmers[2] and A Dahlgren[3]
[1]Centre for Epidemic Interventions Research, Norwegian Institute of Public Health, 0213 Oslo, Norway
[2]Centre for Evidence-Based Medicine, University of Oxford, OX2 6GG, UK
[3]Faculty 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 that can help you avoid being misled by untrustworthy treatment claims. In this essay, we explain how five seemingly logical assumptions about research can be misleading. These assumptions are that:

- a plausible explanation is sufficient,
- association is the same as causation,
- more data is better data,
- a single study is sufficient, or
- fair comparisons are not applicable in practice.

The basis for these concepts is described elsewhere.[1]

Do not assume that a plausible explanation is sufficient

Treatments that should work in theory often do not work in practice or may turn out to be harmful. A plausible explanation of how or why a treatment might work does not prove that it actually does work, or that it is safe. For example, cutting someone to make them bleed (bloodletting) used to be a common treatment for lots of problems. People believed it would rid the body of 'bad humours', which is what they thought made people sick. But bloodletting did not help. It even killed people, including George Washington, the first president of the United States.[2] His doctors drained 40% of his blood to treat a sore throat!

A more recent theory was that operating on blocked tubes (arteries) that carry blood to the brain would stop damage to the brain (strokes). That makes sense, but when that theory was tested in a fair comparison, researchers found not only that it did not help, but that some people died from the surgery.[3]

Even if there is plausible evidence that a treatment works in ways likely to be beneficial, the size of any such treatment effect, and its safety, cannot be predicted. For example, most drugs in a class of heart medicines called beta-blockers have beneficial effects in reducing recurrence of heart attacks; but two drugs in the class – pronethalol and practolol – were taken off the market because of unanticipated side effects.[4] Similarly, it cannot be assumed that a treatment works or does not work based on the type of treatment. For example, it cannot be assumed that all complementary medicines or that all modern medicines do or do not work, or that all vaccines do or do not work. On the other hand, not understanding how a treatment works does not mean that it does not work.

Do not assume that association is the same as causation

The fact that a possible treatment outcome (i.e. a potential benefit or harm) is associated with a treatment does not mean that the treatment caused the outcome. The association or correlation could instead be due to chance or some other underlying factor. For example, people who seek and receive a treatment may be healthier and have better living conditions than those who do not seek and receive the treatment. Therefore, people receiving the treatment might appear to benefit from the treatment, but the difference in outcomes could be because they are healthier and have better living conditions, rather than because of the treatment.

An obvious example of confusing an association with causation would be to assume that going to the

doctor causes people to be sick because going to the doctor is associated with being sick. It is more likely that people went to the doctor because they were sick than that going to the doctor caused them to be sick. Another obvious example would be to assume that eating ice cream causes people to drown because ice cream sales are associated with drowning. A more likely explanation for that association is that when it is hot people eat more ice cream and they also swim more. In this example, hot weather is a confounder – it is associated with the 'treatment' (eating ice cream) and it affects the 'outcome' (the number of people who drown).

A less obvious example of confusing an association with causation was the assumption that hormone replacement therapy (HRT) prevented cardiovascular disease (CVD). For many years, experts and doctors believed that HRT reduced the risk of CVD, based on an association found in studies that compared women who chose to take HRT and some women assigned to HRT experienced an increased risk of CVD. However, large, randomised trials did not show any benefit or an increased risk of CVD in women assigned to HRT. An explanation for this is that socioeconomic status was a confounder in the non-randomised studies. Women of lower socioeconomic status are more likely to have CVD and are less likely to take HRT. So, a reason for the association found in the non-randomised studies was the difference in socioeconomic status between the comparison groups, not the difference in whether they took HRT or not.[5]

Do not assume that more data is better data

Claims that are based on 'big data' (data from large databases) or 'real-world data' (routinely collected data) can be misleading. More data simply gives a more statistically precise estimate of whatever biases there might be in a treatment comparison using routinely collected data. When using routinely collected data, it is only possible to control for confounders that are already known and have been measured. Unfortunately, routinely collected data often do not include sufficient detail to confidently conclude that any association found between a treatment and an outcome means that the treatment caused the outcome.

For example, routinely collected (real-world) data have been used in non-randomised comparisons of different types of coronary artery bypass surgeries. Twelve studies including 34,019 patients used a non-randomised study design that is believed to reduce the risk of bias due to confounders (propensity-score matching).[6] They found that using two internal thoracic arteries compared to using one artery was associated with a lower risk of dying within one year. A more likely explanation is that the association was because of confounders that had not been measured. Using two arteries instead of one increases the complexity and invasiveness of the surgery. It is likely that surgeons tend to reserve this type of surgery for patients perceived as healthier and expected to live longer. This type of bias in allocating patients to different treatments (e.g. based on the individual surgeon's judgement) is very difficult to quantify. The statistics can only be adjusted for the measured confounders.[7] As a further illustration of this problem, a large, randomised trial found little or no difference in survival after 10 years. This contrasts with 14 non-randomised studies using propensity-score matching with 24,123 patients. These found that using two arteries improved survival compared to one artery.[8] This was due to lower survival in patients in randomised trials, who were allocated to the two-artery group, and higher survival in the group allocated to the one-artery group compared to the studies using 'real-world data'.

Describing routinely collected data as 'real-world data' implies that data collected in carefully designed fair comparisons of treatments do not come from the real world. Databases of routinely collected data may indeed include a broader spectrum of people than data collected in fair comparisons of treatments that have narrow eligibility criteria. However, routine collection of data is rarely planned to include the information that is needed to ensure fair comparisons, and randomised trials can be designed to have wide eligibility criteria.

Do not assume that a single study is sufficient

The results of one study considered in isolation can be misleading. A single comparison of treatments rarely provides conclusive evidence; and results are often available from other comparisons of the same treatments. Systematic reviews of all the similar comparisons ('replications') may yield different results from those based on the initial studies, and these should help to provide more reliable and statistically precise estimates of treatment differences. Even so, obtaining reliable estimates from treatment comparisons must always consider that important studies may remain unpublished, incompletely published or inaccessible for other reasons.

Randomised trials of oral rehydration solutions (ORS) for children with diarrhoea provide an example of single comparisons of treatments that did not provide conclusive evidence.[9] Children with diarrhoea can become dehydrated. If they become seriously dehydrated, they can die. For more than

20 years, the World Health Organization (WHO) recommended a standard ORS with a large amount of sugar and salt mixed in water. However, some researchers believed that it might be better to use a smaller amount of sugar and salt (reduced osmolarity). Eleven randomised trials published between 1982 and 2001 compared ORS with reduced osmolarity to the standard solution. A key outcome was the number of children who needed an unscheduled fluid infusion, which indicates that they were becoming seriously dehydrated. The results varied. It was not until the results of all the studies were carefully summarised in a systematic review that it was shown convincingly that a reduced osmolarity solution was substantially more effective than the standard solution. Based on combined results of all 11 studies, the WHO changed its recommendation.

Replication or reproducibility is sometimes used to describe the extent to which similar studies, such as the trials of reduced osmolarity ORS, have similar results. However, these terms are not well defined and can sometimes cause confusion.[10]

Do not assume that fair comparisons are not applicable in practice

Assumptions that fair comparisons of treatments in research are not applicable in practice can be misleading. People may claim that evidence from fair comparisons of treatments cannot be applied to everyday practice. This is likely to be true if there are important differences between the fair comparisons and everyday practice. The effects of treatments are unlikely to differ substantially unless there are compelling reasons for why everyday practice is so different from the fair comparisons that the treatments are unlikely to work in the same way.[11]

Deciding whether there are compelling reasons depends on evidence outside fair comparisons of treatments (for example, basic science research that demonstrates how a treatment causes an outcome) and judgement. Reasons for uncertainty about the applicability of research only become compelling when there is compelling evidence or compelling logical reasons for expecting the effects of a treatment to be substantially different in practice.

For example, human biology tends to be more similar than different across people from different countries, races and ethnicities. So, you would expect medicines to have similar effects most of the time. Thus, it is not necessary to conduct randomised trials of medicines in every country with large samples of people from every race and ethnicity. But there are sometimes important differences. For example, the benefits of lowering elevated blood pressure in reducing strokes and other cardiovascular morbidity and mortality are well established. However, several different types of medicine are used to lower blood pressure and there has been uncertainty about which of these should be used. There has also been uncertainty about whether these medicines worked in the same way in Black people and in non-Black people, particularly for angiotensin-converting enzyme (ACE) inhibitors. This is because ACE inhibitors were found to be less effective for lowering blood pressure in Black people than in non-Black people. For this reason, a randomised trial designed to compare different medicines for lowering blood pressure planned to do a subgroup analysis for Black participants in the trial, which included 33,357 participants (35% Black) in the USA and Canada.[12] The results of this study were largely similar for Blacks and non-Blacks, except for the effect of the ACE inhibitor on strokes. Black participants assigned to the ACE inhibitor were more likely to have a stroke than Black participants assigned to the thiazide diuretic, but this difference was not found in non-Black participants.

Various terms are used to describe the 'applicability' of research, including transferability, generalisability, external validity and relevance. Although these terms have been defined differently, checklists designed to assess these concepts include broadly similar criteria.[13] These include differences between fair comparisons and everyday practice in the characteristics of the people, characteristics of the treatments and characteristics of the context. It is possible to generate long lists of things that could potentially be different. For example, differences in patient characteristics could include differences in age, sex, education, income, race, ethnicity, weight, co-morbidity, genetic markers, astrological sign, baseline risk, etc. To avoid being misled by spurious assumptions about fair comparisons not being relevant, only those factors for which there are compelling reasons for why a treatment is unlikely to work the same way in practice as it did in fair comparisons should be considered when assessing the applicability of research results.

It should be noted that most often the relative treatment effect will be similar for people with different baseline risks. Differences in baseline risk will, however, often lead to differences in the absolute effect of treatment.

Implications

- Do not assume that claims about the effects of treatments based on an explanation of how they

might work are correct if the treatments have not been assessed in systematic reviews of fair comparisons of treatments.
- Do not assume that an outcome associated with a treatment was caused by the treatment unless other reasons for the association have been ruled out in a systematic review of fair comparisons.
- Do not assume that an association between a treatment and an outcome found using 'big data' or 'real-world data' means that the treatment caused the outcome unless other possible reasons for the association have been ruled out.
- The results of single comparisons of treatments can be misleading. Consider all the relevant fair comparisons when making judgements about treatment effects.
- Do not assume that fair comparisons are not applicable because of differences between fair comparisons and everyday practice, unless there are compelling reasons for why treatments would work differently.

Declarations

Competing Interests: None declared.

Funding: 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 conceptualized, reviewed, and edited drafts of this essay. ADO prepared the first draft.

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

References

1. Oxman AD, Chalmers I and Dahlgren A; 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. See http://doi.org/10.5281/zenodo.6611932.
2. Morens DM. Death of a president. *N Engl J Med* 1999; 341: 1845–1849.
3. Powers WJ, Clarke WR, Grubb RL Jr, Videen TO, Adams HP Jr and Derdeyn CP. Extracranial-intracranial bypass surgery for stroke prevention in hemodynamic cerebral ischemia: the Carotid Occlusion Surgery Study randomized trial. *JAMA* 2011; 306: 1983–1992.
4. Furberg CD, Herrington DM and Psaty BM. Are drugs within a class interchangeable? *Lancet* 1999; 354: 1202–1204.
5. Humphrey LL, Chan BK and Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 2002; 137: 273–284.
6. Gaudino M, Di Franco A, Rahouma M, Tam DY, Iannaccone M, Deb S, et al. Unmeasured confounders in observational studies comparing bilateral versus single internal thoracic artery for coronary artery bypass grafting: a meta-analysis. *J Am Heart Assoc* 2018; 7: e008010.
7. Agoritsas T, Merglen A, Shah ND, O'Donnell M and Guyatt GH. Adjusted analyses in studies addressing therapy and harm: users' guides to the medical literature. *JAMA* 2017; 317: 748–759.
8. Gaudino M, Rahouma M, Hameed I, Khan FM, Taggart DP, Flather M, et al. Disagreement between randomized and observational evidence on the use of bilateral internal thoracic artery grafting: a meta-analytic approach. *J Am Heart Assoc* 2019; 8: e014638.
9. Hahn S, Kim S and Garner P. Reduced osmolarity oral rehydration solution for treating dehydration caused by acute diarrhoea in children. *Cochrane Database Syst Rev* 2002: Cd002847.
10. Goodman SN, Fanelli D and Ioannidis JP. What does research reproducibility mean? *Sci Transl Med* 2016; 8: 341ps12.
11. Dans AL, Dans LF, Guyatt GH and Richardson S; Group ftE-BMW. Users' guides to the medical literature XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998; 279: 545–549.
12. Wright JT Jr, Dunn JK, Cutler JA, Davis BR, Cushman WC, Ford CE, et al. Outcomes in hypertensive black and nonblack patients treated with chlorthalidone, amlodipine, and lisinopril. *JAMA* 2005; 293: 1595–1608.
13. Munthe-Kaas H, Nøkleby H and Nguyen L. Systematic mapping of checklists for assessing transferability. *Syst Rev* 2019; 8: 22.