Reliably measured outcomes

From: Key Concepts for assessing claims about treatment effects and making well-informed treatment choices (Version 2022)

2.1e Consider whether outcomes were assessed reliably.

Explanation

Some outcomes are easy to assess, such as births and deaths. Others are more difficult, such as depression or quality of life. For treatment comparisons to be meaningful, outcomes that are meaningful to people should be assessed using methods that have been shown to be reliable.

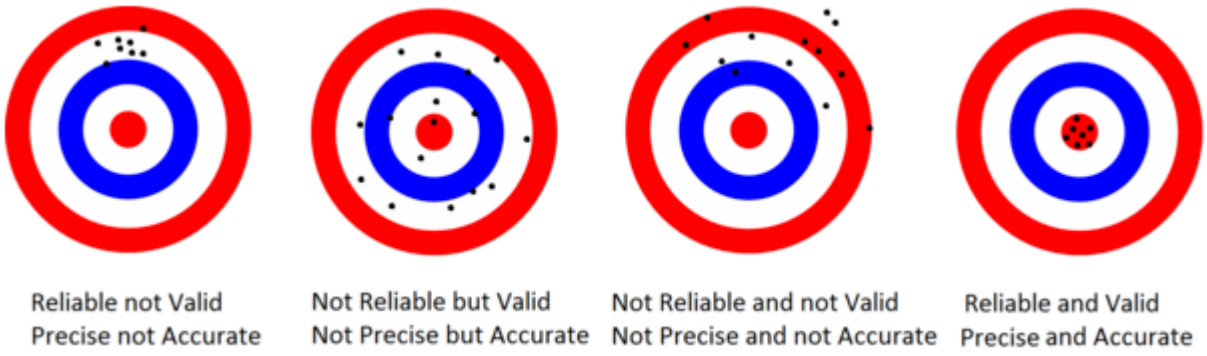
Unreliable outcome measures result in outcome misclassification or measurement error. When misclassification is similar in the groups of people being compared ("non-differential"), this tends to lead to underestimation of effects. For example, a vaccine cannot be expected to protect against infections other than those for which it was developed. So, for example, influenza vaccines are less effective for preventing 'influenza-like' illness (much of which is not caused by influenza viruses) than for preventing influenza that is confirmed by a laboratory test *[Demicheli 2018 (SR)]*. As the proportion of influenza-like illnesses that are caused by influenza viruses decreases, the difference will increase between the effects of vaccines on influenza-like illness and laboratory confirmed influenza.

Basis for this concept

Non-differential (unbiased) misclassification of outcomes has been shown to lead to underestimation of treatment effects in simulations and research studies *[Blackwelder 1991 (RS), Copeland 1977 , De Smedt 2018 , Hunnicutt 2016 (SR), Petersen 2021 (SR), Rogers 2006 (RS), Walraven 2018 (RS)]*. Underestimation of effects increases as the accuracy of the assessment decreases and as the occurrence of the outcome decreases.

Non-randomized studies of the association between treatments (and other factors) and outcomes mention the concept of measurement error in about half of the studies published in top-ranked journals *[Brakenhoff 2018 (SR)]*. An additional problem in non-randomized studies is error in measuring exposure to the treatments being studied and to confounders. Unlike treatment outcome measurement error, exposure measurement error does not always lead to underestimation of effects. Few studies investigate the impact of measurement error, so it is difficult to judge the robustness of the reported effect estimates.

A target is often used to explain the difference between precision or the extent of random errors (sometimes referred to as reliability) and "validity" or systematic errors (sometimes referred to as accuracy) in outcome measures or diagnostic tests (Figure).



For self-reported outcomes, systematic errors can be caused by social desirability bias *[Althubaiti 2016]*. Self-reporting of an outcome (or a treatment, in non-randomized studies) can be influenced by social desirability or approval, especially when anonymity and confidentiality cannot be guaranteed. For example, self-reporting of behaviours such as diet, smoking, sexual behaviours, drug use, or compliance with a prescribed treatment can be influenced by the study participants' perceptions of what the investigators or others view as good or bad behaviour. This can result in over-reporting of "good behaviours" and underreporting of "bad behaviours".

Measuring outcomes that are important to patients often depends on patient-reported outcomes *[Calvert 2013 , Garratt 2002 (SR), Johnston 2021]*. When patient-reported quality of life is reported in randomized trials, the reported effects on quality of life sometimes are not in agreement with the primary outcome measures *[Contopoulos-Ioannidis 2009 (SR)]*.

However, for estimates of the effects of treatments on patient-reported outcomes to be reliable, the patient-reported outcome measures used must be reliable and valid *[Gagnier 2021]*. Outcomes that are measured using an outcome measure that has not been shown to be reliable and valid can result in misleading effect estimates. For example, randomized trials of treatments for schizophrenia that used unpublished outcome measures were more likely to report that a treatment was superior to the comparison (control) treatment compared to trials that had used a published (evaluated) outcome measure *[Marshall 2000 (SR)]*.

Patient-reported outcome measures do not always reflect what is meaningful and important to patients. It is important to ensure that patients understand them and that they capture what is important to them. For example, the McGill Pain Questionnaire is widely used in randomized trials, however it was developed for clinician reporting and never underwent qualitative evaluation with direct patient input. Interviews with patients likely would have revealed difficulties understanding the options for responding to the question "How strong is your pain?" than the ones that are used (1 Mild, 2 Discomforting, 3 Distressing, 4 Horrible, 5 Excruciating) *[Basch 2011]*.

It is also important that patient-reported outcome measures are comprehensive include all aspects of an outcome that are important to patients and relevant (do not include aspects that are unimportant or irrelevant). For example, a systematic review of patient-reported outcome measures for postpartum recovery included 15 outcome measures *[Sultan 2021 (SR)]*. The obstetric-specific outcome measures included between four and 12 aspects ("domains") of outpatient postpartum recovery. They were all missing at least one domain, such as pain, psychosocial distress, sleep, motherhood experience, fatigue, or sexual function. On the other hand, some outcome measures include domains that are unlikely to be relevant in some settings, such as "satisfaction with pollution" and "satisfaction with transportation".

The number of patient-reported outcome measures is growing rapidly, and there are now well over 1,000 systematic reviews of those measures *[COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) 2019]*. Variation in the outcome measures that are reported in different comparisons of the same treatments makes it difficult to synthesise and interpret the results in systematic reviews. A core outcome set is a standardised set of outcomes, agreed by stakeholders, including patients that should be the minimum outcomes measured and reported in all trials in particular health areas *[Matvienko-Sikar 2021 (SR)]*. Core outcome sets are often not used or reported in randomized trials. Greater use of core outcome sets could improve evaluations of treatment effects and systematic reviews. Use of core outcome sets could also reduce the risk of selective outcome reporting (see Concept 2.2b), enhance research transparency, and help to ensure that important outcomes are assessed using reliable outcome measures.

Implications

Be cautious about relying on the results of treatment comparisons if outcomes have not been assessed using methods that have been shown to be reliable.

References

Systematic reviews

Brakenhoff TB, Mitroiu M, Keogh RH, Moons KGM, Groenwold RHH, van Smeden M. Measurement error is often neglected in medical literature: a systematic review. J Clin Epidemiol. 2018;98:89-97. https://doi.org/10.1016/j.jclinepi.2018.02.023

Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JP. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. BMJ. 2009;338:a3006. https://doi.org/10.1136/bmj.a3006

Demicheli V, Jefferson T, Ferroni E, Rivetti A, Di Pietrantonj C. Vaccines for preventing influenza in healthy adults. Cochrane Database Syst Rev. 2018;2(2):Cd001269. https://doi.org/10.1002/14651858.cd001269.pub6

Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ. 2002;324(7351):1417. https://doi.org/10.1136/bmj.324.7351.1417

Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, Lapane KL. Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: a systematic review. Pharmacoepidemiol Drug Saf. 2016;25(12):1343-53. https://doi.org/10.1002/pds.4076

Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. Br J Psychiatry. 2000;176:249-52. https://doi.org/10.1192/bjp.176.3.249

Matvienko-Sikar KL, Fellow HRBR, Avery KSL, Blazeby JP, Devane DP, Dodd SL, et al. Use ofcore outcome sets was low in clinical trials published in major medical journals. J Clin Epidemiol. 2021. https://doi.org/10.1016/j.jclinepi.2021.10.012

Petersen JM, Ranker LR, Barnard-Mayers R, MacLehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. Int J Epidemiol. 2021;50(5):1708-30. https://doi.org/10.1093/ije/dyab061

Sultan P, Sharawi N, Blake L, Ando K, Sultan E, Aghaeepour N, et al. Use of Patient-Reported Outcome Measures to Assess Outpatient Postpartum Recovery: A Systematic Review. JAMA Netw Open. 2021;4(5):e2111600. https://doi.org/10.1001/jamanetworkopen.2021.11600

Research studies

Blackwelder WC, Storsaeter J, Olin P, Hallander HO. Acellular pertussis vaccines. Efficacy and evaluation of clinical case definitions. Am J Dis Child. 1991;145(11):1285-9. https://doi.org/10.1001/archpedi.1991.02160110077024

Rogers WO, Atuguba F, Oduro AR, Hodgson A, Koram KA. Clinical case definitions and malaria vaccine efficacy. J Infect Dis. 2006;193(3):467-73. Clinical Case Definitions and Malaria Vaccine Efficacy

Walraven CV. A comparison of methods to correct for misclassification bias from administrative database diagnostic codes. Int J Epidemiol. 2018;47(2):605-16. https://doi.org/10.1093/ije/dyx253

Other references

Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211-7. https://doi.org/10.2147/jmdh.s104807

Basch E, Abernethy AP, Reeve BB. Assuring the patient centeredness of patient-reported outcomes: content validity in medical product development and comparative effectiveness research. Value Health. 2011;14(8):965-6. https://doi-org.ezproxy.uio.no/10.1016/j.jval.2011.10.002

Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA. 2013;309(8):814-22. https://doi.org/10.1001/jama.2013.879

COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN). COSMIN Database of Systematic Reviews, 2019. Accessed: 22 November 2021. https://www.cosmin.nl/tools/database-systematic-reviews/

Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol. 1977;105(5):488-95. https://doi.org/10.1093/oxfordjournals.aje.a112408

De Smedt T, Merrall E, Macina D, Perez-Vilar S, Andrews N, Bollaerts K. Bias due to differential and non-differential disease- and exposure misclassification in studies of vaccine effectiveness. PLoS One. 2018;13(6):e0199180. https://doi.org/10.1371/journal.pone.0199180

Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021;30(8):2197-218. https://doi.org/10.1007/s11136-021-02822-4

Johnston B, Patrick D, Devji T, Maxwell L, Bingham III C, Beaton D, et al. Patient-reported outcomes. Cochrane Handbook for Systematic Reviews of Interventions version 6,2. 2021. https://training.cochrane.org/handbook/current/chapter-18