Similar comparison groups

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

2.1a Consider whether the people being compared were similar.

Explanation

If people in treatment comparison groups differ in ways other than the treatments being compared, the apparent effects of the treatments might reflect those differences rather than actual treatment effects. Differences in the characteristics of the people in the comparison groups at the beginning of the comparison might result in estimates of treatment effects that appear either larger or smaller than they actually are. A method such as allocating people to different treatments by assigning them random numbers (the equivalent of flipping a coin) is the best way to ensure that the groups being compared are similar in terms of both measured and unmeasured characteristics.

If people are not randomly allocated to treatment comparison groups, differences between the groups other than the treatments may result in estimates of treatment effects appearing larger or smaller than they actually are because of confounders or other differences. For example, patients who are most ill (e.g., have severe pain) may be more likely to be given a new treatment than patients who are less ill. There may appear to be a sharp response to treatment in the most ill patients because of regression to the mean. If they are compared to patients who are less ill and receive an older treatment, the new treatment may appear to be more effective than it actually is compared to the older treatment. Differences in recall ("recall bias") can also lead to over- or under-estimates of effects in case-control and retrospective cohort studies that are based on recollection of exposure to a treatment.

As described in relation to Concept 1.2b, the effect of hormone replacement therapy (HRT) on cardiovascular disease (CVD) is an example of overestimation of a treatment effect in non-randomized studies. For many years experts and doctors believed that HRT reduced the risk of CVD, based on non-randomized studies. But the results of large, randomized trials provided no support for this belief and sometimes suggested an increased risk of CVD in women assigned to HRT. This may be because women of lower socio-economic status are more likely to have CVD and less likely to take HRT. So, a reason for the apparent beneficial effect of HRT on CVD in non-randomized studies is the difference in socioeconomic status between the comparison groups, rather than the difference in whether they took HRT or not *[Humphrey 2002 (SR)]*.

Quinidine is an example of a treatment for which a beneficial effect appeared smaller in non-randomized studies when compared to those in randomized studies. Quinidine was frequently used to treat heart rhythm abnormalities (atrial fibrillation).[1] A systematic review of randomized and non-randomized studies found that the beneficial effect of maintaining a normal heart rhythm was 54% less after three months and 76% less after 12 months in non-randomized studies when compared with randomized studies *[Reimold 1992 (SR)]*. One possible explanation for the apparently smaller effects in the non-randomized studies is that patients with the most symptoms and the highest risk may have been more likely to receive quinidine in the non-randomized studies.

[1] Although quinidine was effective for maintaining a normal heart rhythm, it has been replaced by safer and more effective medicines.

Aspirin is an example of a treatment where a harmful effect appeared larger in non-randomized studies when compared to randomized studies. Randomized studies have shown that low-dose aspirin reduces the risk of stroke in people at high risk (with symptoms and signs of vascular disease) but not in people at low risk. A systematic review of randomized and non-randomized studies found an increased risk of stroke in people at low risk who took aspirin, whereas randomized studies did not find an increased risk *[Hart 2000 (SR)]*. Aspirin use in the non-randomized studies was largely self-selected and it is possible that people who chose to take aspirin had a higher risk of stroke than those who did not, even after statistical adjustment for risk factors that were known and had been measured.

Basis for this concept

Random allocation of people to comparison groups is unbiased with respect to prognosis (characteristics of participants that can predict the course and outcome of a condition) and responsiveness to the treatment. No other way of creating comparison groups has these properties because it cannot be assumed that all factors relevant to prognosis and responsiveness to treatment have been distributed in an unbiased way between comparison groups *[Kleijnen 1997]*. However, when a small number of people are randomly allocated, important differences between comparison groups can occur by chance. Moreover, both randomized studies and non-randomized studies can be misleading for other reasons *[Sterne 2016]*, including those addressed by Key Concepts 2.1b to 2.1g.

Comparisons of the results of randomized and non-randomized studies have found that carefully designed and implemented non-randomized studies and randomized studies sometimes give similar estimates of the effects of treatments *[Anglemyer 2014 (SR), Bun 2020 (SR), Concato 2000 (SR), Golder 2011 (SR), Schwingshackl 2021 (SR)]*. However, non-randomized comparisons of treatments can overestimate effects, underestimate effects, mask effects, or reverse the direction of effects *[Deeks 2003 (SR), Ewald 2020 (SR), Hemkens 2016a (SR), Ioannidis 2001 (SR), Kunz 1998 (SR), Odgaard-Jensen 2011 (SR)]*. It is a paradox that the unpredictability of randomization is the best protection against the unpredictability of the extent and direction of bias in treatment comparisons that are not properly randomized.

To ensure that people in treatment comparison groups are similar, in addition to randomly allocating enough people, it is important to ensure that random allocation is properly implemented. Researchers have investigated the impact of two key elements of random allocation: adequate generation of a random sequence (to ensure that the allocation sequence is unpredictable, and that people are allocated by chance), and concealed allocation (to ensure that the random sequence is properly implemented, and that participation is not influenced by knowing the treatment assignment prior to enrolment in the study. A systematic review combined the data from seven studies that investigated the influence of these and other characteristics of randomized trials on effect estimates *[Savović 2012b (SR)]*. It included 234 meta-analyses containing 1,973 randomized trials. It found that, on average, effects were overestimated in trials with inadequate or unclear (compared with adequate) random-sequence generation and with inadequate or unclear (compared with adequate) allocation concealment. A systematic review of 24 studies found similar results *[Page 2016a (SR)]*. A review of 56 studies that examined associations between 58 different trial characteristics and effect estimates found that allocation concealment, sequence generation, and small sample size were the characteristics most consistently associated with treatment effect estimates *[Dechartres 2016 (SR)]*. However, it is not generally possible to predict the magnitude, or even the direction, of bias in studies with inadequate or unclear random-sequence generation or allocation concealment *[Armijo-Olivo 2015 (SR), Bialy 2014 (SR), Bolvig 2018 (SR), Ginnerup-Nielsen*

2016 (SR), Hartling 2014 (SR), Koletsi 2016 (SR), Odgaard-Jensen 2011 (SR), Saltaji 2018 (SR), Wang 2021 (SR)].

Implications

Be cautious about relying on the results of non-randomized treatment comparisons (for example, if the people being compared chose which treatment they received). Be particularly cautious when you cannot be confident that the characteristics of the comparison groups are similar. If people were *not* randomly allocated to treatment comparison groups, ask if there were important differences between the groups that might have resulted in the estimates of treatment effects appearing either larger or smaller than they actually are.

References

Systematic reviews

Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014(4). https://doi.org//10.1002/14651858.MR000034.pub2

Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open. 2015;5(9):e008562. https://doi.org/10.1136/bmjopen-2015-008562

Bialy L, Vandermeer B, Lacaze-Masmonteil T, Dryden DM, Hartling L. A meta-epidemiological study to examine the association between bias and treatment effects in neonatal trials. Evid Based Child Health. 2014;9(4):1052-9. https://doi.org/10.1002/ebch.1985

Bolvig J, Juhl CB, Boutron I, Tugwell P, Ghogomu EAT, Pardo JP, et al. Some Cochrane risk-of-bias items are not important in osteoarthritis trials: a meta-epidemiological study based on Cochrane reviews. J Clin Epidemiol. 2018;95:128-36. https://doi.org/10.1016/j.jclinepi.2017.11.026

Bun RS, Scheer J, Guillo S, Tubach F, Dechartres A. Meta-analyses frequently pooled different study types together: a meta-epidemiological study. J Clin Epidemiol. 2020;118:18-28. https://doi.org/10.1016/j.jclinepi.2019.10.013

Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342(25):1887-92. https://doi.org/10.1056/nejm200006223422507

Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. J Clin Epidemiol. 2016;77:24-37. https://doi.org/10.1016/j.jclinepi.2016.04.005

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):iii-x, 1-173. https://doi.org/10.3310/hta7270

Ewald H, Ioannidis JPA, Ladanie A, Mc Cord K, Bucher HC, Hemkens LG. Nonrandomized studies using causal-modeling may give different answers than RCTs: a meta-epidemiological study. J Clin Epidemiol. 2020;118:29-41. https://doi.org/10.1016/j.jclinepi.2019.10.012

Ginnerup-Nielsen E, Christensen R, Thorborg K, Tarp S, Henriksen M. Physiotherapy for pain: a meta-epidemiological study of randomised trials. Br J Sports Med. 2016;50(16):965-71. https://doi.org/10.1136/bjsports-2015-095741

Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med. 2011;8(5):e1001026. https://doi.org/10.1371/journal.pmed.1001026

Hart RG, Halperin JL, McBride R, Benavente O, Man-Son-Hing M, Kronmal RA. Aspirin for the primary prevention of stroke and other major vascular events: meta-analysis and hypotheses. Arch Neurol. 2000;57(3):326-32. https://doi.org/10.1001/archneur.57.3.326

Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. PLoS One. 2014;9(2):e88008. https://doi.org/10.1371/journal.pone.0088008

Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. BMJ. 2016a;352:i493. https://doi.org/10.1136/bmj.i493

Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. Ann Intern Med. 2002;137(4):273-84. https://doi.org/10.7326/0003-4819-137-4-200208200-00012

Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286(7):821-30. https://doi.org/10.1001/jama.286.7.821

Koletsi D, Spineli LM, Lempesi E, Pandis N. Risk of bias and magnitude of effect in orthodontic randomized controlled trials: a meta-epidemiological review. Eur J Orthod. 2016;38(3):308-12. https://doi.org/10.1093/ejo/cjv049

Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ. 1998;317(7167):1185-90. https://doi.org/10.1136/bmj.317.7167.1185

Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schunemann H, et al. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev. 2011(4):MR000012. https://doi.org/10.1002/14651858.mr000012.pub3

Page MJ, Higgins JP, Clayton G, Sterne JA, Hróbjartsson A, Savović J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. PLoS One. 2016a;11(7):e0159267. https://doi.org/10.1371/journal.pone.0159267

Reimold SC, Chalmers TC, Berlin JA, Antman EM. Assessment of the efficacy and safety of antiarrhythmic therapy for chronic atrial fibrillation: observations on the role of trial design and implications of drug-related mortality. Am Heart J. 1992;124(4):924-32. https://doi.org/10.1016/0002-8703(92)90974-z

Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, da Costa BR, Flores-Mir C. Impact of Selection Bias on Treatment Effect Size Estimates in Randomized Trials of Oral Health Interventions: A Meta-epidemiological Study. J Dent Res. 2018;97(1):5-13. https://doi.org/10.1177/0022034517725049

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 Intern Med. 2012b;157(6):429-38. https://doi.org/10.7326/0003-4819-157-6-201209180-00537

Wang Z, Alahdab F, Farah M, Seisa M, Firwana M, Rajjoub R, et al. Association of study design features and treatment effects in trials of chronic medical conditions: a meta-epidemiological study. BMJ Evid Based Med. 2021. https://doi.org/10.1136/bmjebm-2021-111667

Research studies

Schwingshackl L, Balduzzi S, Beyerbach J, Bröckelmann N, Werner SS, Zähringer J, et al. Evaluating agreement between bodies of evidence from randomised controlled trials and cohort studies in nutrition research: meta-epidemiological study. BMJ. 2021;374:n1864. https://doi.org/10.1136/bmj.n1864

Other references

Kleijnen J, Gøtzsche P, Kunz RA, Oxman AD, Chalmers I. So what's so special about randomisation? In: Maynard A, Chalmers I, editors. Non-Random Reflections on Health Care Research: On the 25th Anniversary Of Archie Cochrane's Effectiveness and Efficiency. London: BMJ Publishing Group; 1997. p. 93-106.

Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. https://doi.org/10.1136/bmj.i4919