Outcomes counted in the right group

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

2.1g Consider whether people's outcomes were analysed in the group to which they were allocated.

Explanation

Random allocation to treatment comparison groups helps to ensure that people in the comparison groups have similar characteristics before they receive treatment (see Concept 2.1a). However, people sometimes do not receive or take the treatment allocated to them. The characteristics of such people often differ from those who do take the treatments allocated to them. Excluding from the analysis people who did not receive the treatments allocated to them may mean that like is no longer being compared with like. This may lead to an underestimate or an overestimate of treatment differences relative to what would have been the case if everyone had received treatment that had been intended for them.

For example, in a comparison of surgery and drug treatments, people who die while waiting for surgery should be counted in the surgery group, even though they did not receive surgery. This may seem counter-intuitive. But if they are excluded and people who die during the same time in the drug group are not excluded, it will not be a fair comparison.

The New York Health Insurance Plan (HIP) randomized trial of screening for breast cancer provides a striking illustration of how people who comply with a treatment (in this case, screening mammography) may be different from those who do not. The study found similar numbers of deaths after five years among women offered screening and those who were not offered screening (Table) *[Shapiro 1977 (RS)]*. Some women offered screening chose not to be screened. If those women are excluded from the comparison, it appears that there were fewer deaths in the screened group compared to the women who were not offered screening (22 versus 30 per 1,000 women). However, that comparison is misleading because there were important differences between the women offered screening who chose to be screened and those who chose not to be screened. Those differences resulted in almost twice as many deaths among women who chose not to be screened compared to women who chose to be screened (40 versus 22 per 1,000 women).

Table. Total number of deaths after five years in the HIP randomized trial of breast cancer screening*

Comparison group	Group size	Deaths per 1,000 women
Offered screening	31,000	28
Chose to be screened	20,200	22
Chose not to be screened	10,800	40
Not offered screening	31,000	30

* Data from Table 1 in *[Freedman 2004]*.

Basis for this concept

A systematic review of randomized trials published in the top five medical journals reported the results in three ways *[Mostazir 2021 (SR)]*:

1) Including outcomes in all the study participants allocated to each of the treatment comparison groups ("intention-to-treat" analysis)
   In this analysis, study participants who dropped out of the study, did not adhere to the study treatment to which they were allocated, or even took the wrong study treatment, are included in the treatment comparison group to which they were randomly allocated.

2) Only including outcomes in participants who adhered to the trial protocol, including the treatment to which they were allocated ("per-protocol" analysis).
   The aim of this analysis is to answer the question: "What is the effect if participants are fully compliant?" However, because it excludes participants who were not compliant, the treatment comparison groups will no longer be similar if the people who do not comply and are not included in the analysis, as illustrated by the breast cancer screening example above.

3) Using a statistical model to compare people who complied in the "treatment group" to people in the "control group" who would have complied to the study treatment (using the "complier average causal effect" (CACE) method).

On average, the "per-protocol" analyses generated larger estimates of treatment effects than the intention-to-treat analyses. Differences between the two analyses increased with increasing degrees of non-compliance. However, the CACE effect estimates were similar to the intention-to-treat estimates, suggesting that the "per-protocol" analyses likely overestimated the impact of non-compliance on the intention-to-treat effect estimates. In other words, they were biased.

Other systematic reviews have compared the results of randomized trials that included all participants in the analysis (intention-to-treat) to the results of trials comparing the same treatments but after excluding some participants. These reviews have found that, on average, trials that did not report intention-to-treat analyses over-estimated or underestimated treatment effects compared to those that used intention-to-treat analyses *[Abraha 2015 (SR)]* treatment effects *[Armijo-Olivo 2021 (SR)]*. Other reviews were inclusive about the average impact of excluding participants *[Balk 2002 (SR), de Almeida 2019 (SR), Siersma 2007 (RS), van Tulder 2009 (SR)]*, or found that this often resulted in biased estimates of treatment effects, but the extent and direction of bias was unpredictable *[Nüesch 2009 (SR)]*. One systematic review found that many systematic reviews of the effects of treatments include at least one randomized trial that did not report an intention-to-treat analysis, and that those trials were more likely to have "positive" ("statistically significant") findings, industry sponsorship, and authors with conflicts of interest *[Abraha 2017 (SR)]*. All the reviews that compare randomized trials, have a high risk of confounding by other characteristics of the trials. Nonetheless, they support the logical arguments for being cautious about analyses of randomized trials that exclude some participants from the treatment group to which they were allocated.

Implications

Be cautious about relying on the results of treatment comparisons if patients' outcomes have not been counted in the group to which the patients were allocated.

References

Systematic reviews

Abraha I, Cherubini A, Cozzolino F, De Florio R, Luchetta ML, Rimland JM, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. BMJ. 2015;350:h2445. https://doi.org/10.1136/bmj.h2445

Abraha I, Cozzolino F, Orso M, Marchesi M, Germani A, Lombardo G, et al. A systematic review found that deviations from intention-to-treat are common in randomized trials and systematic reviews. J Clin Epidemiol. 2017;84:37-46. https://doi.org/10.1016/j.jclinepi.2016.11.012

Armijo-Olivo S, da Costa BR, Ha C, Saltaji H, Cummings GG, Fuentes J. Are biases related to attrition, missing data, and the use of intention to treat related to the magnitude of treatment effects in physical therapy trials? A meta-epidemiological study. Am J Phys Med Rehabil. 2021. https://doi.org/10.1097/phm.0000000000001837

Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA. 2002;287(22):2973-82. https://doi.org/10.1001/jama.287.22.2973

de Almeida MO, Saragiotto BT, Maher C, Costa LOP. Allocation Concealment and Intention-To-Treat Analysis Do Not Influence the Treatment Effects of Physical Therapy Interventions in Low Back Pain Trials: a Meta-epidemiologic Study. Arch Phys Med Rehabil. 2019;100(7):1359-66. https://doi.org/10.1016/j.apmr.2018.12.036

Mostazir M, Taylor G, Henley WE, Watkins ER, Taylor RS. Per-Protocol analyses produced larger treatment effect sizes than intention to treat: a meta-epidemiological study. J Clin Epidemiol. 2021;138:12-21. https://doi.org/10.1016/j.jclinepi.2021.06.010

Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Bürgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. BMJ. 2009;339:b3244. https://doi.org/10.1136/bmj.b3244

van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine. 2009;34(16):1685-92. https://doi.org/10.1097/brs.0b013e3181ab6a78

Research studies

Shapiro S. Evidence on screening for breast cancer from a randomized trial. Cancer. 1977;39(6 Suppl):2772-82. https://doi.org/10.1002/1097-0142(197706)39:6%3C2772::aid-cncr2820390665%3E3.0.co;2-k

Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. Stat Med. 2007;26(14):2745-58. https://doi.org/10.1002/sim.2752

Other references

Freedman DA, Petitti DB, Robins JM. On the efficacy of screening for breast cancer. Int J Epidemiol. 2004;33(1):43-55. https://doi.org/10.1093/ije/dyg275