Statistical significance

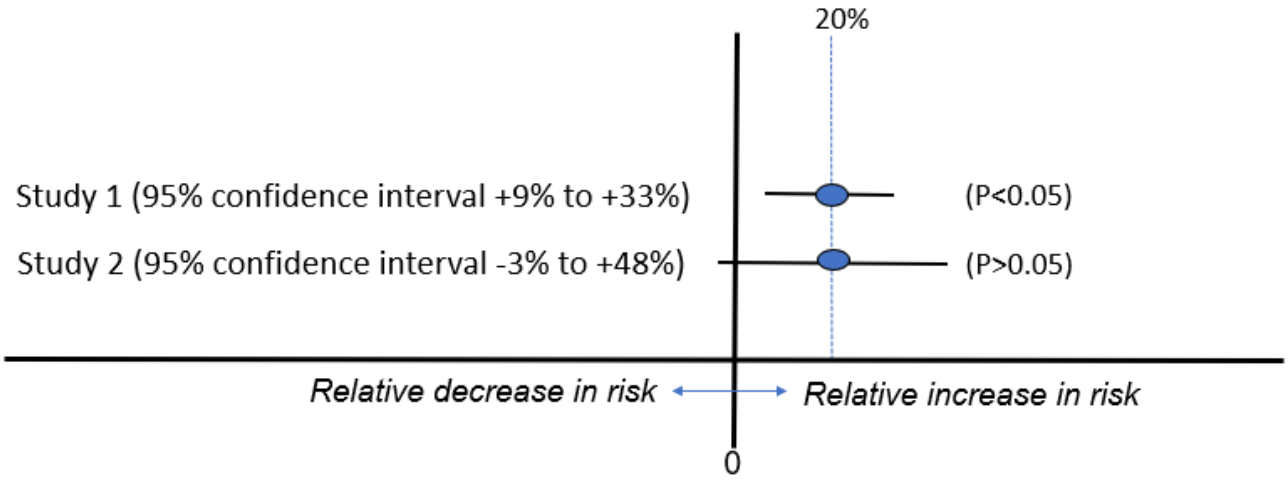
From: Key Concepts for assessing claims about treatment effects and making well-informed treatment choices (Version 2022)

2.4d Be cautious of results reported as "statistically significant" or "non-significant".

Explanation

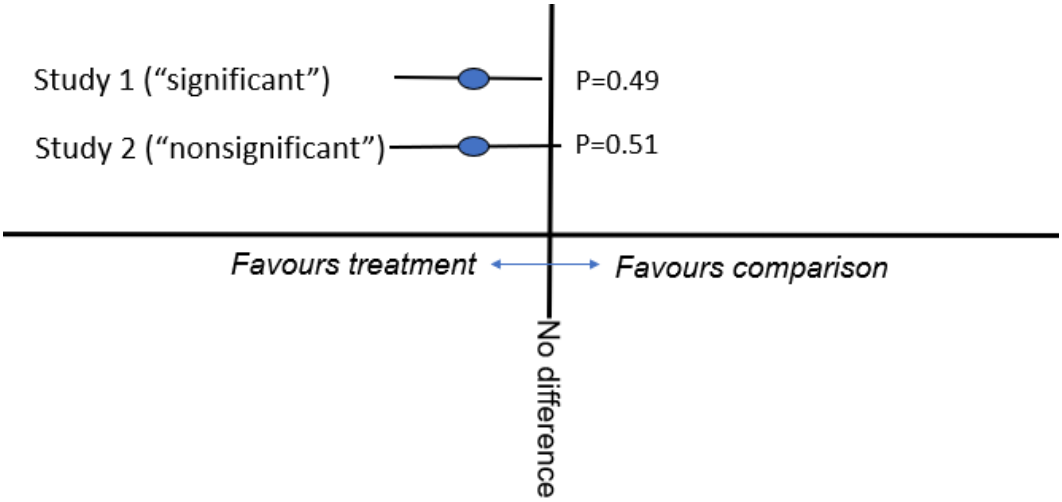
"Statistical significance may be confused with "importance". The cut-off for considering a result as statistically significant is arbitrary, and statistically non-significant results can be either informative (showing that it is very unlikely that a treatment has an important effect) or inconclusive (showing that the relative effects of the treatments compared are uncertain).

For example, two studies of a possible adverse effect of anti-inflammatory drugs on the risk of heart rhythm abnormalities (atrial fibrillation) were reported as having had "statistically nonsignificant" results *[Schmidt 2014]*. The authors of one of the articles concluded that exposure to the drugs was "not associated" with an increased risk and that the results stood in contrast to those from an earlier study with a "statistically significant" result. However, the effect estimates were the same for the two studies: a risk ratio of 1.2 (that is, a 20% relative increase). The earlier study was simply more precise, as indicated by the narrower confidence interval in the figure below. Concluding that the results of the second study showed "no association" was misleading, considering that the confidence interval ranged from a 3% decrease in risk to a 48% increase. It is also misleading to conclude that the results were in contrast with the earlier study that had an identical observed effect. Yet, misleading interpretations like this, which are based on an arbitrary cut-off for "statistical significance" are common.



Basis for this concept

The arbitrariness of a cut-off for "statistical significance is illustrated in the figure below. The results of the two studies are almost identical. Yet, Study 1 is "statistically significant" and Study 2 is "statistically nonsignificant".



In addition to being arbitrary, “significant” and “important” are synonyms, and statistical significance is often confused with importance, especially when “significant” is not prefaced by “statistically”. Statistical significance does not convey any information about the size of the effect. A “statistically significant” effect may or may not be important. Similarly, an observed effect that is “statistically nonsignificant” may or may not be important, and the results may or may not rule out an important effect (see Concept 2.3d).

Systematic reviews have found that the reporting and interpretation of randomized trials with “statistically nonsignificant” findings was frequently inconsistent with the results and biased *[Boutron 2010 (SR)]*, with some authors supporting treatments despite evidence that they might be ineffective or harmful *[Hewitt 2008 (RS)]*, while over half inappropriately interpreted “statistically nonsignificant” results as indicating no effect *[Freiman 1978 (RS), Gates 2019 (RS)]*. On the other hand, reports of findings that were marginally “statistically significant (p-values between 0.01 and 0.10) commonly failed to convey uncertainty when describing the results and often conveyed them as definitively demonstrating an effect” *[Rubinstein 2019 (SR)]*. In addition, results that are “statistically significant” are more likely to be reported in abstracts compared to the corresponding full text *[Assem 2017 (SR), Boutron 2010 (SR), Chavalarias 2016 (SR), Ginsel 2015 (SR), Gøtzsche 2006 (SR)]*, whereas, “statistically nonsignificant results may not be reported at all” (see Concept 2.2b).

Implications

Claims that results were ‘significant’ or ‘non-significant’ usually mean that they were ‘statistically significant’ or ‘statistically non-significant’. This is not the same as ‘important’ or ‘not important’. Do not be misled by such claims.

References

Systematic reviews

Assem Y, Adie S, Tang J, Harris IA. The over-representation of significant p values in abstracts compared to corresponding full texts: A systematic review of surgical randomized trials. Contemp Clin Trials Commun. 2017;7:194-9. https://doi.org/10.1016/j.conctc.2017.07.007

Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA. 2010;303(20):2058-64. https://doi.org/10.1001/jama.2010.651

Chavalarias D, Wallach JD, Li AH, Ioannidis JP. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. JAMA. 2016;315(11):1141-8. https://doi.org/10.1001/jama.2016.1952

Ginsel B, Aggarwal A, Xuan W, Harris I. The distribution of probability values in medical abstracts: an observational study. BMC Res Notes. 2015;8:721. https://doi.org/10.1186/s13104-015-1691-x

Gøtzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. BMJ. 2006;333(7561):231-4. https://doi.org/10.1136/bmj.38895.410451.79

Rubinstein SM, Sigworth EA, Etemad S, Martin RL, Chen Q, Warner JL. Indication of measures of uncertainty for statistical significance in abstracts of published oncology trials: a systematic review and meta-analysis. JAMA Netw Open. 2019;2(12):e1917530. https://doi.org/10.1001/jamanetworkopen.2019.17530

Research studies

Freiman JA, Chalmers TC, Smith H, Jr., Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. N Engl J Med. 1978;299(13):690-4. https://doi.org/10.1056/nejm197809282991304

Gates S, Ealing E. Reporting and interpretation of results from clinical trials that did not claim a treatment difference: survey of four general medical journals. BMJ Open. 2019;9(9):e024785. https://doi.org/10.1136/bmjopen-2018-024785

Hewitt CE, Mitchell N, Torgerson DJ. Listen to the data when results are not significant. BMJ. 2008;336(7634):23-5. https://doi.org/10.1136/bmj.39379.359560.ad

Other references

Schmidt M, Rothman KJ. Mistaken inference caused by reliance on and misinterpretation of a significance test. Int J Cardiol. 2014;177(3):1089-90. https://doi.org/10.1016/j.ijcard.2014.09.205