Risk of bias

Ranstam J, Cook JA

The ideal statistical analysis has two main properties: the result is a fair (or unbiased) assessment of the data with regards to the research question of interest, and furthermore it is sufficiently precise to enable a firm conclusion to be made. When drawing inferences from research findings, how well the analysis attains to this standard should always be considered. Statistical analyses which produce p-values, standard errors, and confidence intervals indicate the potential influence of random errors when generalising the findings in the group of study patients to the entire an underlying population of patients represented by this group. Such random errors arise when sampling from a biologically diverse population and due to other sources of variation (such as measurement errors).

In addition to random error, bias is not uncommon. For example, the readings from a sphygmomanometer for a group of patients, in addition to being very imprecise (having large random errors), can also be systematically off (biased); e.g., they are consistently wrong in that the device produces blood pressure measurements that are on average 4 mmHg too high as well as highly variable. For this reason such devices should be calibrated. When it comes to the result of a research study, systematic errors can occur in various ways and at any stage of the project (e.g. selection, recall, and collider stratification bias)[1]. While the statistical precision to some degree always depends on the sample size, the magnitude of bias is not influenced by the size of the study. Standard statistical approaches do not directly address the potential for such systematic errors. Bias can though often be prevented, or at least minimized, by careful study design, for example by use of random allocation of treatment and or independent outcome assessment. To some extent, the potential for bias can also be explored in the statistical analysis (or at least it can be conducted so as to minimizing it); this is common for observational studies comparing treatments where adjustment for potential prognostic and confounding factors via the use of a propensity score or a casual model is done to minimize selection bias[1].

When critically appraising a scientific report and identifying suboptimal methodology, it is important to recognize that such imperfections or flaws do not *necessarily* lead to a biased and incorrect finding, nor is it usually readily identifiable when the result is indeed biased. It is more appropriate to consider what the risk of the result being biased is. Through looking at the findings of multiple studies with various methodological approaches, those features which are associated with bias can be identified. What these are varies according to the type of study being conducted and the research question being addressed. For example, for a randomised trial of treatments, appropriate generation of the random allocation along with concealment of future allocations, and blinded outcome assessment (as far as is feasible) are the key concerns [2]. Risk of bias tools are available which guide and formalize the critical appraisal of studies [2-4]. As an investigator, it is important to avoid the potential for bias as far as possible, for example, independent assessment of the outcome of an operation is highly desirable for obvious reasons. The direction of any possible remaining bias (i.e. whether it leads to an over- or underestimation of an effect) should be considered when drawing conclusions.

References

1. Cook JA, Ranstam J. Statistical models and confounding adjustment. *BJS* 2017; 104:787-788. DOI: 10.1002/bjs.10245
2. Savović J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 2012; 16(35).
3. P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, *et al.*, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*, 155 (2011), pp. 529-536.
4. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016; 355 doi: https://doi.org/10.1136/bmj.i4919.