Editorial

Systematic reviews, meta-analyses, randomized trials, and observational studies

The aim of well-designed confirmatory randomized clinical trials is to provide reliable answers to research questions, e.g., an estimated effect of a particular treatment, with specified statistical power. Nevertheless, the findings from such trials are always uncertain because of sampling variation. Confidence intervals show the size of this uncertainty. When 2 or more samples are studied to estimate the same treatment effect, the estimates should not be expected to be identical; random differences can be expected. However, including the estimates in a meta-analysis can provide a combined estimate with better precision than the different trials' individual estimates. Not withstanding this, systematic reviews and meta-analyses have recently become very common and receive major criticism (see, e.g., Ioannidis 2016). We will here describe a few issues to be considered by authors and readers pertaining to the usefulness of meta-analyses of observational studies in particular.

First, the PRISMA Statement guidelines have been developed as an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. Compliance with the guidelines improves the reporting of systematic reviews and meta-analyses and this facilitates the editorial evaluation of the report. *Acta Orthopaedica* and many other journals require that a completed PRISMA Statement checklist is included with manuscripts presenting systematic reviews and meta-analyses.

Second, in well-performed experiments, bias is prevented by the study design. Randomization, concealed treatment allocation, and masking are used to avoid selection bias, confounding bias, and information bias. The internal validity of a trial need therefore not be addressed by adjustments in the statistical analysis as is the case with observational studies. Instead, the analysis can focus on precision, for example by stratifying on randomization stratification factors and by adjusting for baseline value when estimating change from baseline. In contrast, observational studies rely entirely on the statistical analysis for validity adjustments. The same statistical methods, such as regression models or ANCOVA, can be used in both experimental and observational studies, which may give the impression that the same analysis is performed in both cases, but the statistical analysis of an observational study is in general more complex and the results are more uncertain. For example, while randomization prevents confounding bias from all factors, validity adjustments can only be performed for known and measured factors.

Furthermore, the statistical adjustment requires assumptions about cause–effect relationships, separating confounders from mediators and colliders. Observational studies are therefore not well suited as confirmatory studies. They are typically exploratory and their findings open to subjective interpretation. The results from meta-analyses of observational studies are therefore also exploratory, and the scientific value may need to be explained by the authors.

Effect measures can be pooled but are not necessarily adjusted for the same factors, and underlying assumptions may be different. A mixture of randomized trials and observational studies is particularly problematic in a meta-analysis (see Faber et al. 2016). One possibility could perhaps be to split the analysis into 2 parts, one for trials and another for observational studies.

Notably, a third problem, related to the second one, is based on the fact that randomized trials and observational studies use, at least partially, different terminology. Several technical terms have clear definitions in randomized trials but no clear interpretation in observational studies. For example, primary and secondary outcomes are parts of a strategy for addressing multiplicity issues in confirmatory trials, but observational studies are exploratory, not confirmatory. The adverse events that it is possible to study in observational studies are usually those that are causally linked to a studied treatment, e.g., complications and side effects. The standard definition of adverse events in a randomized trial is, however, "any untoward medical occurrence temporally, but not necessarily causally, related to the treatment." This information is usually not available in observational studies.

Using trial terminology in reports of observational studies is a poor idea, possibly misleading and unfortunately common, especially in meta-analyses. It may be relevant to emphasize that the guidelines from the ICMJE (the Vancouver group) recommend against non-technical use of technical terms.

A fourth problem is related to the heterogeneity of the effects included in meta-analyses. A statistical fixed-effect model can be used to combine a pooled treatment effect from a number of different study-specific treatment effects. However, if the study-specific treatment effects are more heterogeneous than

could be expected with respect to sampling variation, an analysis strategy based on estimating a common treatment effect may be too simplistic. To estimate an average treatment effect may be more adequate, and this can be performed using a random-effect model. It is important to recognize that the estimates from a fixed-effect and a random-effect model are fundamentally different (Riley et al. 2011), and that the variability of the effects represented by their estimated average can have clinical relevance. The recommended way to evaluate estimates from random-effect models is by constructing prediction intervals for the treatment effect estimates (see, e.g., Higgins et al. 2009).

A further problem with meta-analysis of heterogeneous studies, however, is that the main criterion for choosing between fixed-effect and random-effect models, I^2 (the fraction of variance that is due to heterogeneity rather than sampling variation), tends to be biased in meta-analyses with small numbers of studies (von Hippel 2015).

In summary, a well-performed systematic review and meta-analysis of well-performed confirmatory randomized trials with similar inclusion criteria and endpoints can contribute new and useful evidence-based information. However, the usefulness of meta-analyses of observational studies that are based on different study populations, different data collection procedures, different statistical analysis strategies, different underlying assumptions, and with bias adjustments for different sets of confounding factors can be debated.

Jonas Ranstam
Philippe Wagner
Statistical Editors
email: jonas.ranstam@med.lu.se
email: philippe.wagner@med.lu.se

Faber T, Ravaud P, Riveros C, Perrodeau C, Dechartres A. Meta-analyses including non-randomized studies of therapeutic interventions: a methodological review. BMC Med Res Methodol 2016: 35.

Higgins J P, Thompson S G, Spiegelhalter D J. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc 2009; 172: 137-59.

Ioannidis J P. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 2016; 94: 485-514.

Riley R D, Higgins J P T, Deeks J J. Research methods & reporting: interpretation of random effects meta-analyses. BMJ 2011; 342: d549.

von Hippel P T. The heterogeneity statistic I^2 can be biased in small meta-analyses. BMC Med Res Methodol 2015; 15: 35.