Short Research Review

Performing and evaluating meta-analyses

Jakob Burcharth, MD, PhD, Hans-Christian Pommergaard, MD, PhD, *and* **Jacob Rosenberg, MD, DMSc,** *Herlev, Denmark*

From the Centre for Perioperative Optimization, Department of Surgery, Herlev Hospital, University of Copenhagen, Herlev, Denmark

LITERATURE REVIEWS can be performed as narrative or systematic. The narrative review often is written by experts and generally provides an extensive overview of the literature; however, the narrative review does not provide transparency of the review process. In contrast, systematic reviews include a description of a predefined research question, a systematic search strategy, and a screening and selection strategy that uses predefined inclusion and exclusion criteria; the systematic review evaluates the quality of the included studies. Together, this type of review potentially allows the reader to reproduce the review process and add transparency and objectivity to the work. If some or all data from the included studies in a systematic review also are presented in a quantitative synthesis, it is named a meta-analysis.[1]

Meta-analyses are powerful tools for summarizing knowledge as well as estimating treatment effects with greater precision. Furthermore, meta-analyses are able to identify possible publications bias (eg, when small studies with undesirable findings are not being published).[2] The number of published meta-analyses are increasing dramatically, and more than 18% of all published meta-analyses indexed in MEDLINE were published in the year 2013 (Fig 1). Despite the usefulness of systematic reviews and meta-analyses in answering important clinical questions, these reviews are sometimes controversial and should be interpreted with care, because results can be misleading if the methodology is inappropriate. Criticism of the meta-analytic method includes that the method is based on data that are extracted and integrated from a number of independent studies instead of random sampling of data, which means that the results from a meta-analysis cannot test causality. This article introduces and should help to clarify the basic methodologic principles of performing and evaluating meta-analyses.

PLANNING AND REGISTERING A REVIEW

The Cochrane Handbook of Systematic Reviews of Interventions[2] describes the details of performing the specific steps of systematic reviews and meta-analyses, and the Preferred Reporting Items of Systematic Reviews and Meta-analysis guideline, which consists of a checklist and a 4-phased flowchart, describes the preferred content of each passage when reporting a systematic review and meta-analysis.[3] The Meta-analyses Of Observational Studies in Epidemiology checklist describes how to report meta-analyses of observational studies.[4] Following Preferred Reporting Items of Systematic Reviews and Meta-analysis or Meta-analyses Of Observational Studies in Epidemiology guidelines ensures systematic reporting and transparency in the review process.

When planning a systematic review and meta-analysis, it is advisable to publish a detailed protocol before commencing the review. By registering a protocol prospectively at the PROSPERO (International Prospective Register of Systematic Reviews) webpage,[5] it is possible to prevent multiple reviews addressing the same question. This registration process will also decrease the risk of publication bias in the event of negative results, because registration will bind review authors to the literature search, analysis plan, bias evaluation,

J.R. received lecture fees from Merck and Bard and research funding from Baxter Healthcare, Johnson & Johnson, and Bard.

Accepted for publication August 22, 2014.

Reprint requests: Jakob Burcharth, MD, PhD, Centre for Perioperative Optimization, Department of Surgery, Herlev Hospital, University of Copenhagen, Herlev 2730, Denmark. E-mail: jakobburcharth@gmail.com.

Surgery 2015;157:189-93.

0039-6060/$ - see front matter

http://dx.doi.org/10.1016/j.surg.2014.08.087

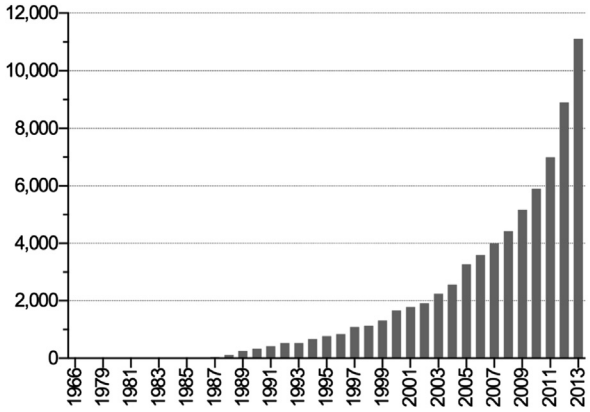

Fig 1. Number of published meta-analyses by year. A total of 52,626 meta-analyses are published in MEDLINE, and of those, 18% were published in the year 2013.

outcome selection, and reported outcomes. The registration is expected to be kept up-to-date, and despite strong encouragement from several opinion-leaders for the registration of reviews at PROSPERO,[6] this process is still voluntary.

The research question of a systematic review and meta-analysis can either be broad or narrow. A broad research question increases the chance of applying the findings to a wider population; however, it also increases the risk of too much variation between the included studies (eg, heterogeneity). A narrow research question can lead to difficulties in finding enough includable studies as well as generalizing the results. In each case, the research question and eligibility criteria of the includable studies should be clinically relevant and may be defined by PICO(S) (ie, Population, Intervention, Comparison, Outcome(s), and Study type).[7]

Risk of bias within studies. The degree of bias within the included studies determines the certainty of which conclusions can be drawn. Just as with the process of screening and study selection, the assessments of risk of bias of the included studies should be performed by at least 2 authors.

Depending on the study design of the included studies, more than 190 different tools have been developed to assess the risk of bias within the studies.[8] In evaluating randomized controlled trials, the Jadad score[9] and the Cochrane risk of bias tool[2] are the most commonly used tools to evaluate risk bias. When nonrandomized studies are being evaluated, the Newcastle-Ottawa scale or the Downs and Black checklist are the recommended tools to evaluate risk bias in these types of studies.[2]

Heterogeneity. Criticism of meta-analyses includes the possibility of inappropriately combined studies that are quite different, which leads to a risk of the results being incorrect reflections of the true effect. Only studies without major bias and with comparable designs, interventions, patients, and measures of outcome should be included and combined.

Studies that are brought together in reviews addressing a specified research question will differ inevitably with a degree of diversity, either because of clinical or methodologic differences among the included studies. In meta-analyses, this difference is termed heterogeneity.[2] Heterogeneity arises when the observed outcome effects determined from the included studies are more different than expected by random chance alone. Clinical differences between the studies can lead to heterogeneity if the outcome is affected by factors that vary across the studies (ie, different patient characteristics or different study interventions). Moreover, methodologic differences can lead to heterogeneity if the included studies are of different design. Assessment of heterogeneity in meta-analyses is crucial, because the results otherwise may be misleading.

It can be argued that some degree of heterogeneity always will exist in meta-analyses, whether or not it can be detected by statistical tests. This determination depends on how heterogeneity is measured and quantified, because heterogeneity can be determined in several different ways. Visually, a lacking overlap in the Forest plot of the horizontal lines representing confidence intervals of the included studies will indicate some degree of heterogeneity (Fig 2).[10] The RevMan program includes measures automatically of the heterogeneity in the Forest plots that determine whether there is a greater spread of the results between the studies than due strictly to chance. One of these heterogeneity measures is the I^2 (also called inconsistency).[2] The I^2 quantifies the degree of variability among the included studies by the use of the χ^2 and the degrees of freedom (dependent on the number of included studies in the meta-analysis) from the pooled estimate. The I^2 can range from 0 to 100% and is easily interpretable since 0% indicates no heterogeneity, and 100% indicates complete heterogeneity. Heterogeneity is generally considered high when $I^2 > 50\%$ and as being substantial when $I^2 > 75\%$.[2] Before reporting a meta-analysis, heterogeneity needs to be investigated and preferably be explained to determine if data can be combined and presented reliably. Normally the pooled estimate should not be reported if $I^2 > 75\%$; however, because no consistent rules exist regarding this