Meta-analyses: what they can and cannot do

Alain J. Nordmann[a], Benjamin Kasenda[a], Matthias Briel[a,b]

[a] Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Switzerland
[b] Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

Summary

Meta-analyses overcome the limitation of small sample sizes or rare outcomes by pooling results from a number of individual studies to generate a single best estimate. As long as a meta-analysis is not limited by poor quality of included trials, unexplainable heterogeneity and/or reporting bias of individual trials, meta-analyses can be instrumental in reliably demonstrating benefit or harm of an intervention when results of individual randomised controlled trials are conflicting or inconclusive. Therefore meta-analyses should be conducted as part of a systematic review, i.e., a systematic approach to answer a focused clinical question. Important features of a systematic review are a comprehensive, reproducible search for primary studies, selection of studies using clear and transparent eligibility criteria, standardised critical appraisal of studies for quality, and investigation of heterogeneity among included studies. Cumulative meta-analysis may prevent delays in the introduction of effective treatments and may allow for early detection of harmful effects of interventions. As opposed to meta-analysis based on aggregate study data, individual patient data meta-analyses offer the advantage to use standardised criteria across trials and reliably investigate subgroup effects of interventions. Network meta-analysis allows the integration of data from direct and indirect comparisons in order to compare multiple treatments in a comprehensive analysis and determine the best treatment among several options.

We conclude that meta-analysis has become a popular, versatile, and powerful tool. If rigorously conducted as part of a systematic review, it is essential for evidence-based decision making in clinical practice as well as on the health policy level.

Key words: meta-analysis; systematic review; bias; randomised controlled trials; evidence-based health care

Introduction

The historical roots of meta-analysis date back to the 17th century when it was supposed in astronomy that the combination of data might be preferable to any individual choice [1]. In the medical field, the statistician Karl Pearson was probably the first to describe formal techniques for combination of data from different studies in 1904 when he examined the preventive effect of serum inoculations against enteric fever [2]. The problem that "any of the groups…are far too small to allow of any definite opinion being informed at all, having regard to the size of the probable error involved", is still one of the most important reasons to conduct meta-analyses today. For many years, however, these techniques were rarely used in medicine as opposed to psychology and educational research where the synthesis of study results enjoyed a growing popularity. Hence, it came to no surprise that a psychologist, Gene Glass, coined the term 'meta-analysis' in 1976 [3]. Three years later the British physician and epidemiologist Sir Archie Cochrane pointed out that people who want to make informed decisions about health care do not have ready access to reliable reviews of the available evidence [4]. Still in the 1970's he initiated a project to systematically identify all controlled clinical trials in the area of pregnancy and
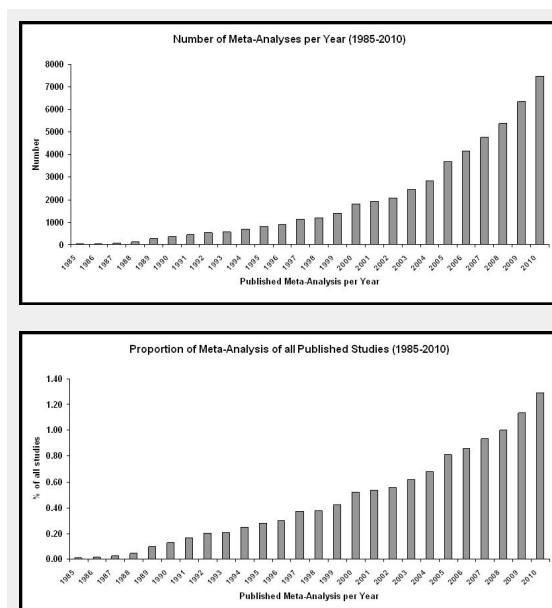

Figure 1

A: Increase in number of meta-analyses in MEDLINE (meta-analysis[mh] OR meta-analysis[pt] OR meta-anal*[tw]) from 1985 to 2010.
B: Increase in proportion of meta-analyses of all published studies in MEDLINE from 1985 to 2010.

obstetrics. From 1974 to 1985 more than 3,500 articles of controlled clinical trials were gathered in a registry and eventually summarised in about 600 systematic reviews. Around the same time Mulrow showed empirically the great potential for error in literature reviews that were not undertaken systematically (so called 'narrative reviews') [5]. As a consequence the necessity of systematic reviews in medical research was increasingly acknowledged. The foundation of *The Cochrane Collaboration* early in the 1990's, an international network of health care professionals who prepare and regularly update systematic reviews (so called 'Cochrane Reviews'), boosted the conduct of meta-analyses in all areas of health care [6]. Meta-analysis has become the most highly cited publication type [7] and its use is still increasing, both in absolute numbers as well as in proportion of all published studies (fig. 1A and 1B).

Aims of this article

In the following paper we will present a typical example from primary care to illustrate how physicians can utilise meta-analyses in general practice. We will describe the differences between narrative and systematic reviews, explain what a meta-analyis is and how it works, acknowledge the strengths but also limitations of meta-analyses, and briefly present advanced tools of meta-analyses such as individual patient data (IPD) and network meta-analysis.

Case vignette

You are the primary care physician of a 65-year old female architect who has been diagnosed with arterial hypertension and dyslipidemia 5 years ago. Both cardiovascular risk factors are well-controlled with amlodipine 5 mg and atorvastatin 10 mg daily. During routine consultation the patient mentions that she is taking 1,000 mg of calcium daily to reduce her risk of experiencing an osteoporotic fracture and asks for your opinion about this prophylactic treatment. How do you advise the patient concerning the continuation of her calcium supplementation?

Osteoporosis is a major cause of morbidity and mortality in elderly people. In a meta-analysis of randomised controlled trials evaluating the effect of calcium on osteoporotic fractures, calcium supplementation was associated with a 12% relative risk reduction in fractures of all types [8]. Consequently, calcium supplements are commonly used by people over the age of 50 to reduce the risk of osteoporotic fractures. However, a five year randomised controlled trial including 732 healthy postmenopausal women reported possible increases in rates of myocardial infarction and cardiovascular events in women allocated to calcium [9], whereas other randomised controlled trials did not report such a harmful association [10, 11].

The modern medical literature is full of examples where study outcomes about a specific intervention seem to contradict each other, leaving the treating physician in a dilemma not knowing whether he should advocate or discourage use of a specific intervention. In situations such as ours, where calcium supplementation may prevent osteoporotic fractures but may increase the risk for myocardial infarction, meta-analysis as part of a systematic review can prove helpful in solving this dilemma.

What is the difference between a narrative and a systematic review?

Narrative reviews are written by experts and qualitatively summarise evidence on a more or less broad topic. They typically use informal, subjective methods to collect and interpret studies, and tend to selectively cite literature that reinforces preconceived notions (table 1) [12]. In contrast, systematic reviews include a comprehensive, reproducible search for primary studies on a focused clinical question, selection of studies using clear and transparent eligibility criteria, critical appraisal of studies for quality, and often a quantitative synthesis of results (meta-analysis) according to a pre-determined and explicit method [13, 14]. A detailed description of further (sub-) types of reviews is given by Grant & Booth [15].

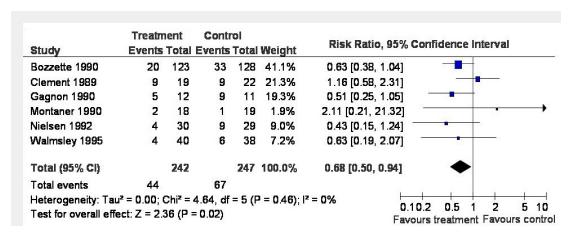## What is a meta-analysis and how does it work?

Figure 2

Forest plot of randomised controlled trials comparing the effect of steroids to placebo/usual care for pneumocystis jirovecii pneumonia on mortality.

The centre of the squares and the horizontal lines correspond to the relative risk or risk ratio (RR) and 95% confidence intervals (CIs). The area of the squares is proportional to the weight each trial contributes to the meta-analysis. The diamond at the bottom of the graph represents the summary RR and its 95% CI indicating a reduction of 32% (95% CI 6 to 50%) in the risk of death when adjunctive corticosteroids are used in pneumocystistis jiroveci pneumonia. The solid vertical line corresponds to no effect of treatment (RR 1.0). The RR, 95% CI and weights are also given in tabular form (adapted from [48] Briel M, Bucher HC, Boscacci R, Furrer H: Adjunctive corticosteroids for Pneumocystis jiroveci pneumonia in patients with HIV-infection. Cochrane Database Syst Rev. 2006 Jul 19;3:CD006150. Reprinted with the permission from John Wiley and Sons).

Table 1: Comparison of narrative and systematic reviews.

Feature	Narrative review	Systematic review
Question	Not explicit/broad	Focused
Search strategy	Not specified	Reproducible
Selection criteria	Absent	Clearly defined
Methodological appraisal of primary studies	Possible/No	Yes
Synthesis of results	Uncommon/qualitative	Quantitative (meta-analysis)

A meta-analysis is the statistical pooling of results from several studies to generate a summary estimate of effects (e.g. from a treatment or a diagnostic tool). Since simple pooling of study results ignoring their precision would yield misleading summary estimates, meta-analysis uses a process of computing weighted averages [16]. This can be accomplished by means of a random effect model or a fixed effect model. Both models can be used to pool a variety of effect measures (discrete and continuous): relative risks, odds ratios, risk differences, p-values, differences in means, sensitivity, specificity, likelihood ratios, etc. The fixed effect model assumes that the studies included in the meta-analysis estimate the same underlying 'true' effect that is fixed, and that the observed differences across studies are due to random error (chance) [14]. The random effect model, on the other hand, assumes that the studies included in the meta-analysis are only a random sample of a theoretical universe of all possible studies on a given research question, and that the effects for the individual studies vary around some overall average effect. Random effects models incorporate two sources of variability: within-study (random error) and between-study variability (heterogeneity). The random effect model is often preferred since it better reflects reality (studies rarely come in identical copies) and usually provides a more conservative estimate with a wider confidence interval.

Ideally, a meta-analysis should be performed as part of a systematic review, but sometimes meta-analyses are done without an initial systematic review and sometimes systematic reviews summarise results only qualitatively and not quantitatively due to considerable differences across studies (see heterogeneity of study results). In meta-analysis, typically, more precise results (larger studies, more events) are assigned more weight in the computation of averages. As any other study, meta-analyses should be conducted according to a pre-specified analysis/research plan.

Reviewers should routinely check for heterogeneity among studies considering the similarity of point estimates, the extent of overlap of confidence intervals (CIs), and statistical criteria such as tests of heterogeneity and the I^2 statistic. I^2 is a measure of inconsistency/heterogeneity among studies in a meta-analysis that can be calculated and compared across meta-analyses of different sizes, of different types of study, and using different types of outcome data [17]. The value of I^2 is reported in %. A low I^2 means that there is little variability between studies that cannot be explained by chance. Although a naive categorisation of values for I^2 is not appropriate for all circumstances, Higgins et al. tentatively assigned adjectives of low, moderate, and high inconsistency to I^2 values of 25%, 50%, and 75%. Whenever heterogeneity is detected in a meta-analysis reviewers need to explore and explain it using methods such as subgroup analysis and meta-regression [18]. Several software packages (e.g., Review Manager, Stata, SAS, R) can perform both fixed and random effect meta-analysis and provide an I^2 statistic. Results of a meta-analysis are usually presented in a Forest plot (fig. 2).

Possible sources to identify meta-analyses

For clinicians, there are several options to search for available meta-analyses on a specific topic in the medical literature. The two databases that are most likely best accessible are MEDLINE and the Cochrane library. In MEDLINE (www.ncbi.nlm.nih.gov) clinicians can use the "clinical queries" function or the search can be restricted by using "meta-analysis" as a limitation under "type of article". In the Cochrane Library (www.thecochranelibrary.com) searches on a specific topic can be restricted to the Cochrane Database of Systematic Reviews and the Database of Abstracts of Reviews of Effects (DARE) to identify systematic reviews and meta-analyses by using the application "Advanced search".

The strengths of meta-analyses

Meta-analyses overcome the limitation of small sample sizes or rare outcomes by pooling results from a number of individual studies and thus increase the statistical power to study effects of interest. Hence, the probability of perceived "negative" results is reduced, and undue delays in the introduction of effective treatments into clinical practice may be avoided. Meta-analysis increases the precision in estimating effects compared to individual trials. They can also contribute to the generalisability of study results. Whereas the findings of a particular study may be limited to the characteristics of this study's population, similar studies' effects in various populations argue for a higher generalisabilty of results. On the other hand, observation of differences in results across various studies (heterogeneity) may allow identification of subgroups (groups of patients defined by certain characteristics e.g., sex or high age) where a specific intervention proves to be particularly beneficial or harmful.
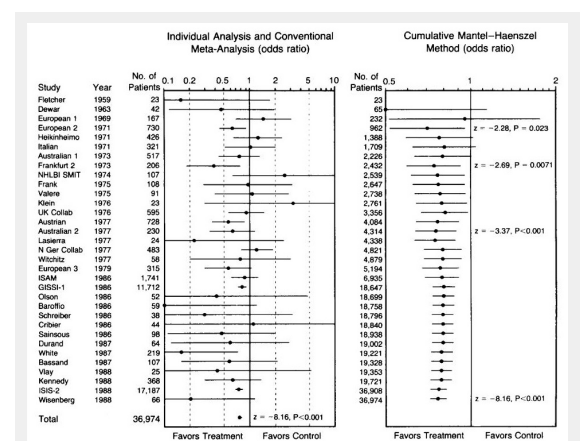

Figure 3

Conventional and cumulative meta-analyses of 33 trials with intravenous streptokinase for acute myocardial infarction. The odds ratios and 95% confidence intervals for an effect of treatment on mortality are shown on a logarithmic scale (from [19], Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992;327:248–54. Reprinted with the permission from the Massachusetts Medical Society "MMS").

A technique called cumulative meta-analysis allows identifying the point in time when the treatment effect of a particular intervention reaches certain levels of statistical significance. Cumulative meta-analysis is the repeated performance of a meta-analysis whenever new trials become available for inclusion. An often quoted example for the potential benefit of cumulative meta-analysis is the publication by Lau et al. demonstrating the reduction in mortality after an acute myocardial infarction thanks to the use of streptokinase [19]. Would a cumulative meta-analysis including every newly published trial evaluating this research question have been performed, a reduction in mortality in patients treated with streptokinase would already have been established around 1977 when over 4,000 patients had been studied yielding a p-value <0.001 (fig. 3). However, in the absence of such an analysis 32,000 more patients had been randomised over the following 11 years causing many myocardial infarction patients to die unnecessarily. In addition to these studies having been unethical, all the money spent to conduct the 25 unnecessary trials could have been spent on more crucial research questions lacking funding. For fairness reasons, we must bear in mind, however, that some statistically significant meta-analyses have later been contradicted by large randomised controlled trials [20]. Realising that results from meta-analyses may not always be trustworthy led to research about various ways of how bias may be introduced and the methods to detect the presence of such biases. Meta-analysis has hereby become a useful tool for empirical "meta-epidemiological" research [21–23]. We will address limitations of meta-analysis in a separate section later.

Cumulative meta-analysis may also prove useful to monitor the safety of a specific intervention. An illustrative example is the cumulative meta-analysis of Jueni et al. about the cardiovascular risk of the cyclo-oxygenase 2 inhibitor rofecoxib [24]. Based on the cumulative meta-analysis rofecoxib should have been withdrawn in the year 2000 and not only 4 years later, since the cumulative meta-analysis proved the harmful effect of rofecoxib leading to an increase in the incidence of myocardial infarction as early as in the year 2000.

So far, we have presented two examples where meta-analyses were successful in clearly demonstrating benefit or harm of an intervention. Another important purpose of a systematic review and meta-analysis is to highlight a potential lack of adequate evidence in particular areas identifying hereby the need for further studies. For instance, a recent meta-analysis of 3 small trials evaluating the effect of statins in patients with dementia [25] concluded that there was insufficient evidence to recommend statins for the treatment of dementia and requested more trial data to answer this question.

Advanced tools: individual patient data (IPD) and network meta-analysis

A particularly powerful method to investigate potential subgroup effects is IPD meta-analysis which involves obtaining individual information or "raw data" on all patients included in each of the trials. This can be a resource intensive and time-consuming effort. In contrast to conventional meta-analyses based on aggregate data from publications, IPD meta-analyses have the following advantages when elucidating possible subgroup differences: (1) All comparisons between subgroups are within study rather than between studies; (2) using individual patient characteristics rather than summary characteristics of patients included in a study prevents misleading inferences due to an ecological fallacy (i.e., the relation with patient averages across trials may not be the same as the relation for patients within trials) [26]. A systematic review and IPD meta-analysis, for example, adressed the impact of high vs low positive end-expiratory pressures (PEEPs) in three randomised trials that enrolled 2,299 adult patients with severe acute lung injury requiring mechanical ventilation [27]. This IPD meta-analysis tested a small number of subgroup hypotheses and found that in patients with severe disease (labeled acute respiratory distress syndrome) higher PEEP is associated with a reduction in hospital mortality and shorter time to unassistant breathing. In patients with mild disease, results suggested that a higher PEEP strategy does not convey benefits and may even be harmful.

Further advantages of IPD meta-analyses include the use of standardised definitions and analyses across studies, accurate ascertainment of all relevant data, and adjustment for variations in individual patient prognosis at baseline.

Most industry trials aiming at obtaining licenses for new products use current or older treatments as comparators. Hence, head-to-head comparisons of new treatments that would be most useful to clinical practice are not available. In an attempt to compare multiple treatments in a comprehensive analysis and determine the best treatment among several options, a method called network meta-analysis (synonyms: multiple-treatments or mixed-treatment comparisons meta-analysis) allows the integration of data from direct (when treatments are compared within a randomised trial) and indirect comparisons (when treatments are compared between trials by combining results on how effective they are compared with a common comparator treatment) [28]. A network meta-analysis of 117 RCTs recently evaluated the efficacy and acceptability of 12 new-generation antidepressants and found that "sertraline might be the best choice when starting treatment for moderate to severe major depression in adults because it has the most favourable balance between benefits, acceptability, and acquisition costs" [29]. The validity of indirect and mixed treatment comparisons depends on certain basic assumptions that are similar to but more complex than assumptions underlying standard meta-analysis [30, 31]. Commonly these assumptions are not statistically verifiable and one has to rely on expert clinical and epidemiological judgment, e.g. when assessing inconsistencies among eligible trials.

Limitations of meta-analyses

Inappropriate meta-analyses can either lead to false negative or false positive results. Meta-analyses yielding false-positive results have the potential of delaying the conduct of large, "definite" trials. Some meta-analyses of small trials were subsequently contradicted by findings of a large randomised controlled trial, e.g., investigations of magnesium for mortality reduction after myocardial infarction, ni-

trates for mortality reduction in myocardial infarction, aspirin for reduction of pregnancy-induced hypertension, or albumin for mortality reduction in the critically ill. How can this happen? There are several mechanisms that may lead to unreliable results of meta-analyses. In the following section we will address 3 important factors that need to be considered when critically evaluating a meta-analysis: inadequate quality of included trials, heterogeneity of study results, and metabias.

Inadequate quality of included trials

Ideally, meta-analyses on therapeutic questions only include high quality randomised controlled trials. For questions about other clinical domains (e.g., prognosis or diagnostic accuracy) different study types (e.g., cohort or cross-sectional studies) may be most suitable. Whatever the question, even well performed meta-analyses cannot correct for the poor quality of trials included in a meta-analysis (also known as "garbage in – garbage out"). The quality of the included studies will always inherently limit the strength of inference one can draw from a meta-analytic summary estimate. Important quality components of randomised trials are concealed treatment allocation (i.e., the impossibility to purposely allocate a specific intervention to a particular patient, for example by central web-based or phone randomisation), blinding of participants, study personnel, outcome assessors and data analysts, small extent and full description of losses to follow-up, and the performance of an intention-to-treat analysis. It remains paramount for all meta-analyses that the methodological quality of included studies is assessed in a standardised manner [32]. In case of inclusion of poor and high quality trials, sensitivity analyses offer the advantage to evaluate the robustness of the results of a meta-analysis by comparing pooled results of high quality to pooled results of poor quality trials. Results of a meta-analysis including individual trials of mixed quality will be more trusted when results are confirmed in the separate pooled analysis of high quality trials alone. Meta-analyses of observational studies for therapeutic questions should be viewed with great caution because they may provide very precise but spurious results due to confounding and selection bias [33].
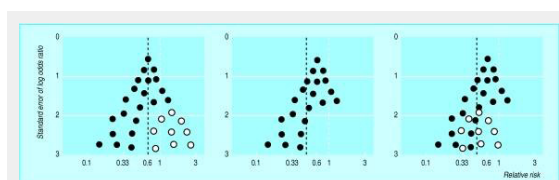

Figure 4

Hypothetical funnel plots: left, symmetrical plot in absence of bias (open circles are smaller studies showing no beneficial effects); centre, asymmetrical plot in presence of publication bias (smaller studies showing no beneficial effects are missing); right, asymmetrical plot in presence of bias due to low methodological quality of smaller studies (open circles are small studies of inadequate quality whose results are biased towards larger effects). Dotted black lines are pooled summary estimates. Pooled estimates exaggerate treatment effects in presence of bias (from Sterne et al. Investigating and dealing with publication and other biases in meta-analysis. BMJ. 2001;323:101–5. Reprinted with the permission from BMJ Publishing Group Ltd.)

Heterogeneity of study results

Meta-analysis of controlled trials is based on the assumption that each included trial provides an unbiased estimate of the effect of an intervention, with the variability of between study results' being attributed to random variation. Ideally, results of individual studies included in a meta-analysis are not heterogeneous. Sometimes, however, a Forest plot will reveal that point estimates of individual studies vary substantially, 95% CIs do not overlap, and I^2 is large. Should that be the case, calculation of a combined effect size from trials may be inappropriate, and, at least, reasons for such heterogenity need to be explored by using meta-regression or stratification [26]. Explanations may lie in the population (e.g., disease severity), the interventions (e.g., doses, co-interventions), the outcomes (e.g., duration of follow-up), the setting (e.g., geographical area, private practice vs hospital), or the study methods (e.g., randomised trials with higher and lower risk of bias). If the latter is true, authors should consider focusing on effect estimates from studies with lower risk of bias. If one of the other categories provides the explanation, authors should offer different estimates across patient groups, interventions, outcomes, or settings. For example, the Forest plot of a meta-analysis of trials of BCG vaccination for the prevention of tuberculosis [34] clearly demonstrated important differences between the effect of the intervention according to the area where BCG vaccination was performed with a larger benefit in warmer than in colder areas. In such situations, it makes more sense to quantify the effect of BCG vaccination for warmer and colder areas separately than to calculate an overall effect estimate. If the observed heterogeneity of results in a meta-analysis can not be sufficiently explained, any reader should be sceptical about the validity of a single pooled estimate.

Reporting bias and other forms of metabias

Much of what we know about bias relates to methodological quality of individual studies. However, there is another form of bias that goes beyond the individual study. Such bias that concerns the available body of evidence on a specific topic rather than an individual study may be called "metabias" [35]. One of the most important metabiases is reporting bias. Only about 50% of randomised trials ultimately reach publication in a journal indexed in a major electronic database and thus become easily identifiable for systematic reviews [36]. What is particularly worrysome is the fact that differences in the dissemination of research findings is not a random process. *Publication bias* describes the phenomenon that the nature and direction of results of an individual study influence publication or non-publication of a study. For example, Turner et al. obtained reviews from the Food and Drug Administration (FDA) for studies of 12 antidepressant agents involving 12564 patients, and conducted a systematic literature search to identify matching publications [37]. According to the published literature (51 trials), it appeared that 94% of the trials conducted were positive. By contrast, the FDA analysis (74 trials) showed that only 51% were positive. Separate meta-analyses of the FDA and journal data sets showed that the increase in effect size ranged from 11 to 69% for individu-

al drugs and was 32% overall. Hence, publication bias may lead to an exaggeration of a treatment effect.

The best way to minimise publication bias is by conducting a comprehensive systematic literature search looking for published trials in several electronic databases (such as Medline, Embase, Cochrane Library, Web of Science, etc), published abstracts by searching recent conference proceedings pertaining to the topic, and unpublished trials by searching trial registries (e.g., www.clinicaltrial.gov) and contacting experts and drug companies in the field. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement requires authors of meta-analyses to describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies), search dates, and details of the electronic search strategy such that it could be repeated [38, 39].

Whether information from scientific abstracts should be incorporated in a meta-analysis is a matter of debate. We suggest including information from scientific abstracts in meta-analyses in an effort to minimise publication bias although this bears the risk of including preliminary, non-peer-reviewed results. In addition we recommend the performance of a sensitivity analysis comparing treatment effects with and without inclusion of data from scientific abstracts in order to assess the robustness of study results.

Using language restrictions to e.g. articles in English may introduce language bias, and selective reporting of some, but not all measured patient-important outcomes leads to outcome reporting bias.

Any reader will by now probably have acknowledged the problem of reporting bias, but will be asking himself: How can I detect reporting bias when certain studies or outcomes have never been published? Funnel plots are probably the most widely used method to evaluate the presence or absence of reporting bias [40]. What is a funnel plot? A funnel plot is a scatter plot of the effect estimates from individual studies against some measure of each study's size or precision. Often the standard error of the effect estimate is chosen as the measure of study size and plotted on the vertical axis with a reversed scale that places the larger, more powerful studies towards the top. Effect estimates from smaller studies should scatter more widely at the bottom, with the spread narrowing among larger studies. In the absence of bias and between study heterogeneity, the scatter will be due to sampling variation alone and the plot will resemble a symmetrical inverted funnel. An asymmetric funnel plot may be an indicator for reporting bias (fig. 4). It is very important, however, to understand that there may be other reasons for funnel plot asymmetry such as methodological limitations in smaller studies that may yield biased overestimates of effects, or the choice of a more restrictive (and thus more responsive) population, or simply chance (for further details about examining and interpreting funnel plot asymmetry see Sterne JA et al. [41]).

An illustrative example of how publication bias may lead to misleading conclusions in meta-analysis is the case of a meta-analysis reporting a reduction in mortality in patients with acute myocardial infarction treated with intravenous magnesium [42], a finding that was subsequently contradicted by the large ISIS-4 trial [43]. Inspection of the cor-

responding funnel plot reveals that selective non-publication of negative trials seems to be a likely explanation for the discrepant findings of ISIS-4 and the magnesium trial meta-analysis [44].

Recently, further types of metabias have been described. Trials that were stopped early for benefit can lead to substantial overestimates of summary effects in meta-analysis if (1) one or more trials were stopped early for benefit after a small number of events (e.g. <200); (2) the difference in treatment effects between trials stopped early for benefit and non-stopped trials is large (e.g., a ratio of RRs <0.7); and (3) the stopped early trials have substantial weight in the meta-analysis (e.g., >20%) [45]. Another empirical study suggested that single centre RCTs (versus multi-centre on the same topic) would be more prone to overestimate treatment effects and thereby leading to inaccurate summary estimates in meta-analyses [23].

Back to our case

After listing the pros and caveats about meta-analysis, let's go back to our 65-year old architect who is taking 1,000 mg calcium daily to reduce her risk for an osteoporotic fracture. After a brief electronic search using the search terms "calcium" and "osteoporosis", and specifying "meta-analysis" as study type under limits in MEDLINE, we identify two systematic reviews and meta-analyses by Bolland et al. on calcium supplements with or without vitamin D and cardiovascular events [46, 47]. Summarising the results of 11 good quality RCTs, the first meta-analysis from 2010 found a pooled RR for myocardial infarction in patients taking versus patients not taking calcium supplementation of 1.27 (95% CI, 1.01 to 1.59) with no evidence of heterogeneity or inconsistency across trials (test for heterogeneity p = 0.96; I^2 = 0%) [46]. In the second meta-analysis from 2011, calcium or calcium with vitamin D increased the risk for myocardial infarction (RR, 1.24; 95% CI, 1.07 to 1.45) and for the composite of myocardial infarction and stroke (RR, 1.15; 95% CI, 1.03 to 1.27) [47]. Alerted by these results, we can now proceed to discuss the potential harms from taking 1,000 mg calcium daily with our patient and weigh it against the reduction in fracture risk that can be expected from calcium supplementation (number needed to treat (NNT) for 5 years of 48) [8]. Based on the meta-analysis by Bolland et al. the number needed to harm (NNH) with calcium for five years to cause a myocardial infarction is 69 [46]. Thanks to these meta-analyses we are now able to provide our patient with the necessary information for making an informed shared decision weighing the risks and benefits of calcium supplementation.

Conclusion

Meta-analysis has become a popular, versatile, and powerful tool in systematically summarising available evidence. In addition to providing a precise estimate of the overall treatment effect in large study populations, meta-analysis may allow early detection of beneficial or harmful treatment effects where individual studies fail to provide reliable treatment estimates. We want to stress that meta-analyses should be performed in the framework of systematic

reviews ensuring standardised assessment of the methodological quality of included studies, appropriate examination of heterogeneity accross studies, and investigations about the completeness of the identified evidence. Rigorously conducted systematic reviews and meta-analyses are essential for evidence-based decision making in clinical practice as well as on the health policy level.

Correspondence: Matthias Briel, MD, MSc, Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Hebelstrasse 10, CH-4031 Basel, Switzerland,
brielm[at]uhbs.ch

References

1 Plackett RL. Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. Biometrika. 1958;45:130–5.

2 Pearson K. Report on certain enteric fever inoculation statistics. Br Med J. 1904;3:1243–6.

3 Glass GV. Primary, secondary and meta-analysis of research. Educ Res. 1976;5:3–8.

4 Cochrane AL. 1931–1971: a critical review, with particular reference to the medical profession. In: Medicines for the Year 2000; London: Office of Health Economics 1979.

5 Mulrow CD. The medical review article: state of the science. Ann Intern Med. 1987;106:485–8.

6 Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. JAMA. 1995;274:1935–8.

7 Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. JAMA. 2005;293:2362–6.

8 Tang BM, Eslick GD, Nowson C, Smith C, Bensoussan A. Use of calcium or calcium in combination with vitamin D supplementation to prevent fractures and bone loss in people aged 50 years and older: a meta-analysis. Lancet. 2007;370:657–66.

9 Bolland MJ, Barber PA, Doughty RN, et al. Vascular events in healthy older women receiving calcium supplementation: randomised controlled trial. BMJ. 2008;336:262–6.

10 Prince RL, Devine A, Dhaliwal SS, Dick IM. Effects of calcium supplementation on clinical fracture and bone structure: results of a 5-year, double-blind, placebo-controlled trial in elderly women. Arch Intern Med. 2006;166:869–75.

11 Baron JA, Beach M, Mandel JS, et al. Calcium supplements for the prevention of colorectal adenomas. Calcium Polyp Prevention Study Group. N Engl J Med. 1999;340:101–7.

12 McAlister FA, Clark HD, van WC, et al. The medical review article revisited: has the science improved? Ann Intern Med. 1999;131:947–51.

13 Glasziou P, Irwig L, Brain C, Colditz G. Systematic reviews in health care. A practical guide. Cambridge: Cambridge University Press 2001.

14 Higgins JPT, Green Se. Cochrane handbook of Systematic Reviews of interventions 2011. www.cochrane-handbook.org (Accessed October 10, 2011) 2011.

15 Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J. 2009;26:91–108.

16 Bravata DM, Olkin I. Simple pooling versus combining in meta-analysis. Eval Health Prof. 2001;24:218–30.

17 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.

18 Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. Stat Med. 2002;21:1503–11.

19 Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992;327:248–54.

20 LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med. 1997;337:536–42.

21 Briel M, Lane M, Montori VM, et al. Stopping randomized trials early for benefit: a protocol of the Study Of Trial Policy Of Interim Truncation-2 (STOPIT-2). Trials. 2009;10:49.

22 Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336:601–5.

23 Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med. 2011;155:39–51.

24 Juni P, Nartey L, Reichenbach S, et al. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. Lancet. 2004;364:2021–9.

25 McGuinness B, O'Hare J, Craig D, et al. Statins for the treatment of dementia. Cochrane Database Syst Rev 2010;CD007514.

26 Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002;21:1559–73.

27 Briel M, Meade M, Mercat A, et al. Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and meta-analysis. JAMA. 2010;303:865–73.

28 Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005;331:897–900.

29 Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet. 2009;373:746–58.

30 Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. Health Technol Assess. 2005;9:1-iv.

31 Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ. 2009;338:b1147.

32 Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

33 Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286:821–30.

34 Colditz GA, Brewer TF, Berkey CS, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. JAMA. 1994;271:698–702.

35 Goodman S, Dickersin K. Metabias: a challenge for comparative effectiveness research. Ann Intern Med. 2011;155:61–2.

36 Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA. 1990;263:1385–9.

37 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008;358:252–60.

38 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151:264–9, W64.

39 Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6:e1000100.

40 Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–34.

41 Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.

42 Yusuf S, Teo K, Woods K. Intravenous magnesium in acute myocardial infarction. An effective, safe, simple, and inexpensive intervention. Circulation. 1993;87:2043–6.

43 ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. Lancet. 1995;345:669–85.

44 Egger M, Smith GD. Misleading meta-analysis. BMJ. 1995;310:752–4.

45 Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA. 2010;303:1180–7.

46 Bolland MJ, Avenell A, Baron JA, et al. Effect of calcium supplements on risk of myocardial infarction and cardiovascular events: meta-analysis. BMJ. 2010;341:c3691.

47 Bolland MJ, Grey A, Avenell A, Gamble GD, Reid IR. Calcium supplements with or without vitamin D and risk of cardiovascular events: reanalysis of the Women's Health Initiative limited access dataset and meta-analysis. BMJ. 2011;342:d2040.

48 Briel M, Bucher HC, Boscacci R, Furrer H. Adjunctive corticosteroids for Pneumocystis jiroveci pneumonia in patients with HIV-infection. Cochrane Database Syst Rev 2006;3:CD006150.
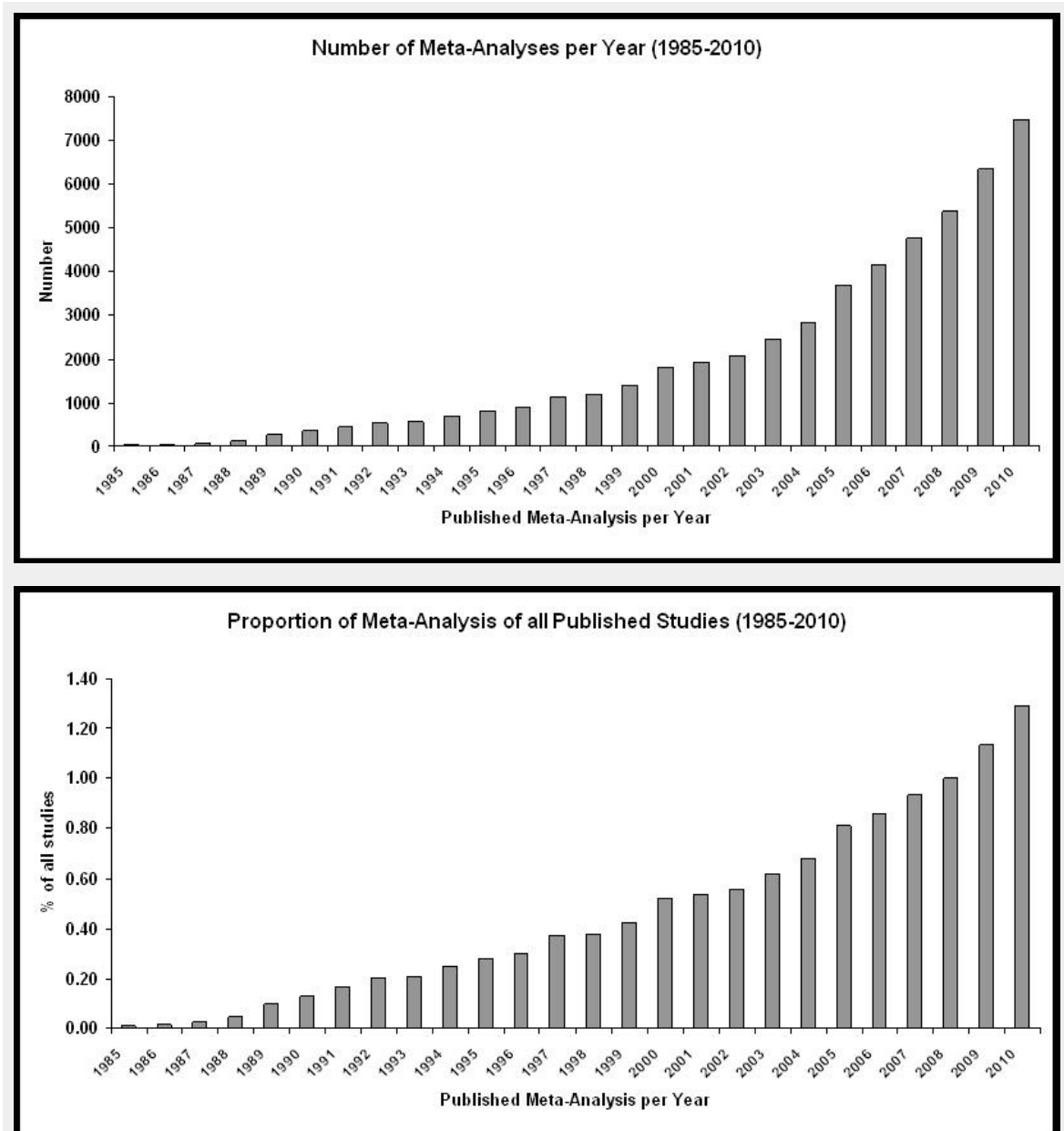
Figures (large format)

Figure 1

A: Increase in number of meta-analyses in MEDLINE (meta-analysis[mh] OR meta-analysis[pt] OR meta-anal*[tw]) from 1985 to 2010.
B: Increase in proportion of meta-analyses of all published studies in MEDLINE from 1985 to 2010.
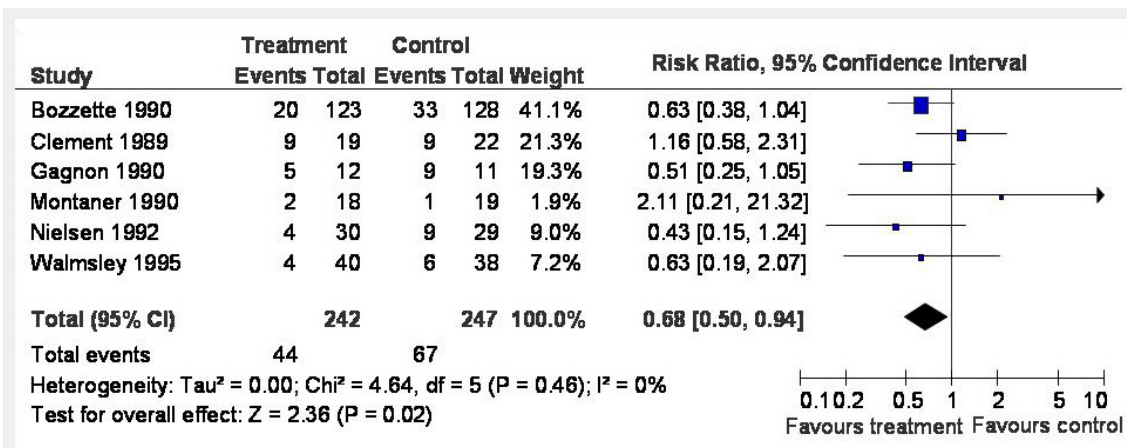
Figure 2

Forest plot of randomised controlled trials comparing the effect of steroids to placebo/usual care for pneumocystis jirovecii pneumonia on mortality.

The centre of the squares and the horizontal lines correspond to the relative risk or risk ratio (RR) and 95% confidence intervals (CIs). The area of the squares is proportional to the weight each trial contributes to the meta-analysis. The diamond at the bottom of the graph represents the summary RR and its 95% CI indicating a reduction of 32% (95% CI 6 to 50%) in the risk of death when adjunctive corticosteroids are used in pneumocystistis jiroveci pneumonia. The solid vertical line corresponds to no effect of treatment (RR 1.0). The RR, 95% CI and weights are also given in tabular form (adapted from [48] Briel M, Bucher HC, Boscacci R, Furrer H: Adjunctive corticosteroids for Pneumocystis jiroveci pneumonia in patients with HIV-infection. Cochrane Database Syst Rev. 2006 Jul 19;3:CD006150. Reprinted with the permission from John Wiley and Sons).
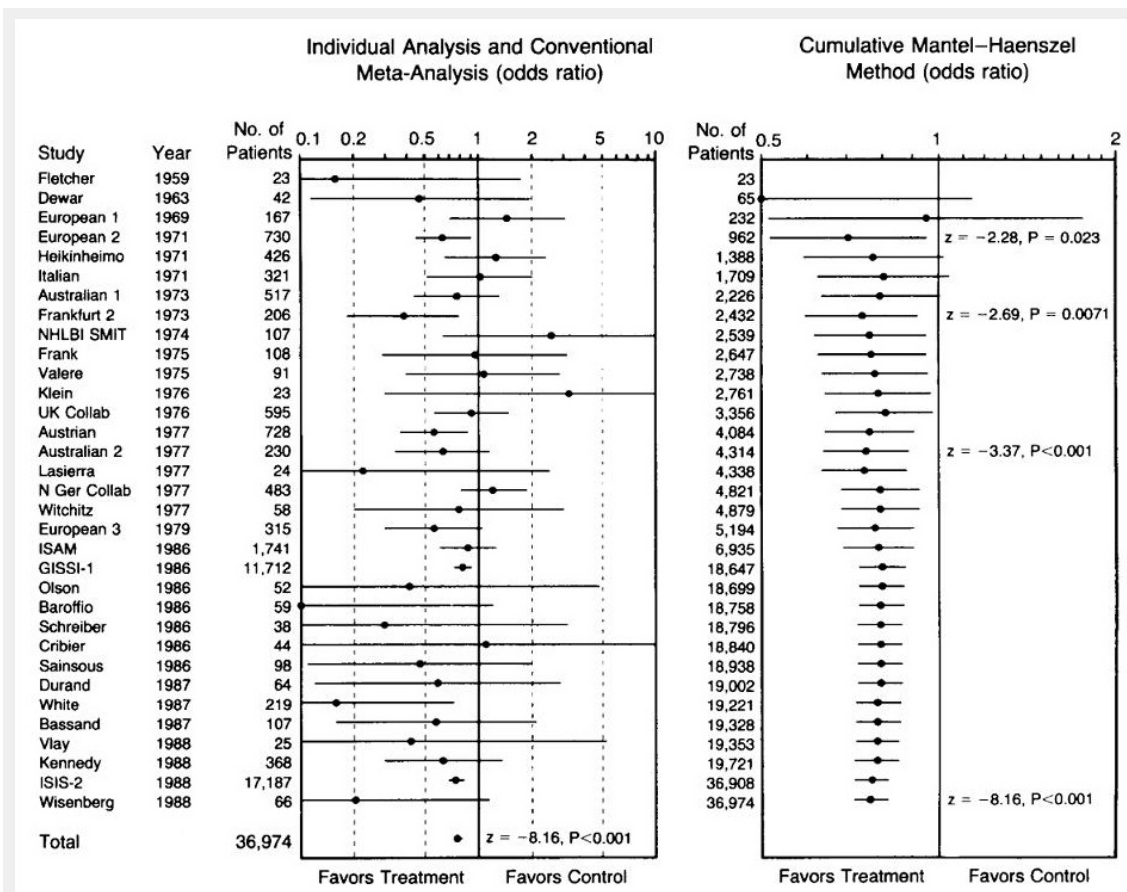

Figure 3

Conventional and cumulative meta-analyses of 33 trials with intravenous streptokinase for acute myocardial infarction. The odds ratios and 95% confidence intervals for an effect of treatment on mortality are shown on a logarithmic scale (from [19], Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992;327:248–54. Reprinted with the permission from the Massachusetts Medical Society "MMS").
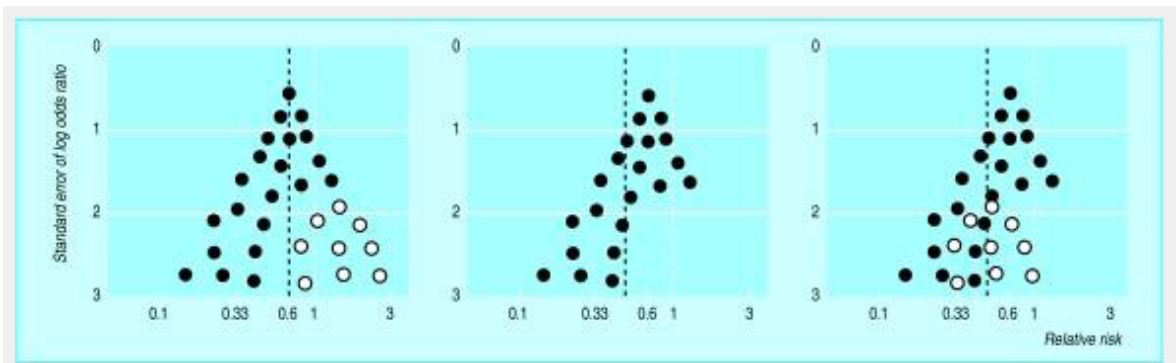
Figure 4

Hypothetical funnel plots: left, symmetrical plot in absence of bias (open circles are smaller studies showing no beneficial effects); centre, asymmetrical plot in presence of publication bias (smaller studies showing no beneficial effects are missing); right, asymmetrical plot in presence of bias due to low methodological quality of smaller studies (open circles are small studies of inadequate quality whose results are biased towards larger effects). Dotted black lines are pooled summary estimates and dotted white lines are null effects (RR 1.0). Pooled estimates exaggerate treatment effects in presence of bias (from Sterne et al. Investigating and dealing with publication and other biases in meta-analysis. BMJ. 2001;323:101–5. Reprinted with the permission from BMJ Publishing Group Ltd.)