Association Between Analytic Strategy and Estimates of Treatment Outcomes in Meta-analyses

Agnes Dechartres, MD, PhD; Douglas G. Altman, DSc; Ludovic Trinquart, PhD; Isabelle Boutron, MD, PhD; Philippe Ravaud, MD, PhD

IMPORANCE A persistent dilemma when performing meta-analyses is whether all available trials should be included in the meta-analysis.

OBJECTIVES To compare treatment outcomes estimated by meta-analysis of all trials and several alternative analytic strategies: single most precise trial (ie, trial with the narrowest confidence interval), meta-analysis restricted to the 25% largest trials, limit meta-analysis (a meta-analysis model adjusted for small-study effect), and meta-analysis restricted to trials at low overall risk of bias.

DATA SOURCES One hundred sixty-three meta-analyses published between 2008 and 2010 in high-impact-factor journals and between 2011 and 2013 in the Cochrane Database of Systematic Reviews: 92 (705 randomized clinical trials [RCTs]) with subjective outcomes and 71 (535 RCTs) with objective outcomes.

DATA SYNTHESIS For each meta-analysis, the difference in treatment outcomes between meta-analysis of all trials and each alternative strategy, expressed as a ratio of odds ratios (ROR), was assessed considering the dependency between strategies. A difference greater than 30% was considered substantial. RORs were combined by random-effects meta-analysis models to obtain an average difference across the sample. An ROR greater than 1 indicates larger treatment outcomes with meta-analysis of all trials. Subjective and objective outcomes were analyzed separately.

RESULTS Treatment outcomes were larger in the meta-analysis of all trials than in the single most precise trial (combined ROR, 1.13 [95% CI, 1.07-1.19]) for subjective outcomes and 1.03 (95% CI, 1.01-1.05) for objective outcomes). The difference in treatment outcomes between these strategies was substantial in 47 of 92 (51%) meta-analyses of subjective outcomes (meta-analysis of all trials showing larger outcomes in 40/47) and in 28 of 71 (39%) meta-analyses of objective outcomes (meta-analysis of all trials showing larger outcomes in 21/28). The combined ROR for subjective and objective outcomes was, respectively, 1.08 (95% CI, 1.04-1.13) and 1.03 (95% CI, 1.00-1.06) when comparing meta-analysis of all trials and meta-analysis of the 25% largest trials, 1.17 (95% CI, 1.11-1.22) and 1.13 (95% CI, 0.82-1.55) when comparing meta-analysis of all trials and limit meta-analysis, and 0.94 (95% CI, 0.86-1.04) and 1.03 (95% CI, 1.00-1.06) when comparing meta-analysis of all trials and meta-analysis restricted to trials at low risk of bias.

CONCLUSIONS AND RELEVANCE Estimation of treatment outcomes in meta-analyses differs depending on the strategy used. This instability in findings can result in major alterations in the conclusions derived from the analysis and underlines the need for systematic sensitivity analyses.


Copyright 2014 American Medical Association. All rights reserved.
Meta-analyses of randomized clinical trials (RCTs) are generally considered to provide among the best evidence of efficacy of medical interventions. They should be conducted as part of a systematic review, a scientifically rigorous approach that identifies, selects, and appraises all relevant studies. Which trials to combine in a meta-analysis remains a persistent dilemma. Meta-analysis of all trials may produce a precise but biased estimate. Thus, the Cochrane Collaboration recommends restricting meta-analyses to trials at low risk of bias, which may result in imprecise estimation of treatment outcomes, or stratifying meta-analyses according to risk of bias. A recent study showed that these recommendations are seldom followed, with only 11% of systematic reviews considering assessment of risk of bias in meta-analyses.

Meta-analysis results can also be affected by small-study effect, defined as the tendency for small trials to show larger treatment outcomes than large trials. A recent study found that this tendency concerned small trials but also moderate-sized trials both when considering trial absolute sample size (eg, fewer than 1000 patients vs more than 1000 patients) and relative sample size (eg, first 3 quarters of sample size within the meta-analysis vs quarter 4 with the largest trials). These results raise the question of whether meta-analyses should be restricted to larger trials (or even to the largest trial). Some authors recently proposed another way to deal with small-study effect with meta-analysis models adjusted for small-study effect. This approach, called “limit meta-analysis,” predicts treatment outcome for a trial of infinite size within a meta-analysis.

In this study, we aimed to compare treatment outcomes estimated by meta-analysis of all trials and several alternative strategies for analysis: single most precise trial, meta-analysis restricted to the largest trials, limit meta-analysis, and meta-analysis restricted to trials at low risk of bias.

Methods

Data Sources

We used combined data from 3 independent collections of meta-analyses of RCTs assessing therapeutic interventions with binary outcomes. The first 2 collections were previously assembled for published meta-epidemiologic studies. Details of the search strategy and selection for these collections of meta-analyses are described elsewhere. Briefly, the first collection included 48 meta-analyses (421 RCTs) published in the 10 leading journals of each medical subject category of the Journal Citation Reports between July 2008 and January 2009 and January and June 2010 or in issue 4 of the Cochrane Database of Systematic Reviews (Issue 4, 2008). We obtained reports of all component trials from included meta-analyses. The second collection included 45 meta-analyses (314 RCTs) published in the Cochrane Database of Systematic Reviews between January and July 2011. The third collection included 70 meta-analyses (505 RCTs) published in the Cochrane Database of Systematic Reviews between April 2012 and March 2013, combining data for 3 RCTs or more. Details of the search strategy and selection of the 3 collections are summarized in eTable 1 and eFigure 1 in the Supplement.

Review and approval of the study by an institutional review board or ethics committee were not applicable because this study, based on published meta-analyses of RCTs, did not directly involve human participants.

Data Collected

As part of the previous meta-epidemiologic studies, for each RCT we extracted data on general characteristics, definition of outcome, results (ie, number of events in each group and number of patients randomized), and assessed risk of bias. Data were extracted from the individual reports of RCTs in the first collection and directly from the Cochrane reviews in the second collection. For the third collection, 2 reviewers independently extracted data from the Cochrane reviews. Disagreements were resolved by discussion with a third reviewer to reach a consensus.

Risk of bias was assessed by the risk of bias tool of the Cochrane Collaboration. Blinding and incomplete outcome data domains were assessed at the outcome level and thus corresponded to the outcome assessed in the meta-analysis. For the first collection, we relied on RCT reports and rated each domain as having low, high, or unclear risk of bias according to the definitions summarized in eTable 2 in the Supplement, following the recommendations of the Cochrane Collaboration. For the second and third collections, we relied on the risk of bias assessment by the review authors. For each RCT, we summarized risk of bias across domains to obtain an overall risk of bias according to the recommendations of the Cochrane Collaboration. The overall risk of bias was classified as low if all key domains were at low risk of bias; as high if at least 1 key domain was at high risk of bias; or as unclear if at least 1 key domain was at unclear risk of bias in the absence of high risk. We considered sequence generation, allocation concealment, blinding, and incomplete outcome data as key domains. We did not consider the domains “selective outcome reporting” and “other risk of bias” because these 2 domains are difficult to assess, particularly for selective outcome reporting when the protocol is not available, which is common.

Classification of Outcomes

We classified outcomes as subjective or objective according to the definitions proposed by Savovíc et al. We considered objective outcomes as all-cause mortality, other objectively assessed outcomes (ie, pregnancy, live births, laboratory outcomes), or outcomes objectively measured but potentially influenced by clinician or patient judgment (eg, hospitalizations, total dropouts or withdrawals, cesarean delivery, assisted delivery, additional treatments administered). We considered subjective outcomes as all other outcomes (ie, patient-reported outcomes, clinician-assessed outcomes, cause-specific mortality). Outcomes were classified independently by 2 reviewers. All disagreements were resolved by discussion to reach consensus.
Data Analysis

Estimation of Treatment Outcome With Different Strategies for Analysis

We estimated treatment outcomes as odds ratios (ORs). Outcome events were recoded so that an OR less than 1 indicated a beneficial association with the experimental intervention.

For each meta-analysis, we estimated treatment outcomes from analytic strategies. Strategy 1 was meta-analysis of all trials. Strategy 2 was the single most precise trial (defined as the trial with the narrowest confidence interval for treatment effect). Strategy 3 was meta-analysis restricted to the largest trials. This strategy involved performing a conventional meta-analysis model but combining data from only the largest trials, excluding smaller trials. We defined the largest trials as those having the largest 25% of sample size within a meta-analysis (ie, those in the fourth quarter of sample size) because a recent meta-epidemiologic study showed larger treatment outcomes for trials in the first three-quarters of sample size than those in the fourth quarter of sample size. Strategy 4 was the meta-analysis described by Rücker et al,11,12 a meta-analysis model including all trials and adjusted for small-study effect. The principle of this method is to predict treatment outcome for a trial of infinite size (ie, a trial that has a treatment outcome with an associated standard error of zero). The method is close to that described by Moreno et al.10,17 Strategy 5 was meta-analysis restricted to trials at low overall risk of bias according to the Cochrane risk of bias tool.

Treatment outcomes were combined across RCTs with use of DerSimonian and Laird random-effects models.18 When appropriate, we used a continuity correction to deal with zero cell counts in 1 group only.19 Heterogeneity across RCTs was assessed by the P statistic.

Comparison of Treatment Outcomes Between Meta-analysis of All Trials and Alternative Strategies

We compared treatment outcomes from meta-analysis of all trials to each alternative strategy with the 2-step meta-epidemiologic approach described by Sterne et al. For each comparison, we applied the following methods. In a first step, for each meta-analysis, we estimated a ratio of odds ratios (ROR), that is, the ratio of the OR for the alternative strategy to the OR for the meta-analysis of all trials. An ROR greater than 1 indicates a larger estimated treatment outcome for meta-analysis of all trials than the alternative strategy. We considered a substantial difference in treatment outcomes between meta-analysis of all trials and the alternative strategy when ROR was outside the range 0.77 to 1.30, indicating a relative difference in treatment outcomes of more than 30% between the strategies. The variance for each log ROR was estimated considering the dependence between the ORs from meta-analysis of all trials and from the alternative strategy: it was derived analytically when there was a single trial for the alternative strategy (ie, single most precise trial, single trial at low risk of bias, single trial in quarter 4 of sample size) or was estimated by the bootstrap method (999 simulations) when there were 2 or more trials for the alternative strategy. Then, in a second step, we estimated a combined ROR across meta-analyses using a random-effects meta-analysis model, which can be interpreted as an average ROR. Heterogeneity of RORs across meta-analyses was assessed by the F statistic and the Cochran Q χ2 test.

Because some previous meta-epidemiologic studies have suggested that the influence of certain trial-level characteristics depended on the type of outcome (ie, subjective vs objective),16,21 we separately analyzed subjective and objective outcomes.

The performance of the limit meta-analysis may be poor when the meta-analysis includes few trials.10,11 As a consequence, we performed a sensitivity analysis including only meta-analyses of 10 trials or more following the rule of thumb used in the area of small-study effect testing.

Exploration of Differences in Treatment Outcomes by Risk of Bias

Because of the results for the comparison of treatment outcomes between meta-analysis of all trials and meta-analysis restricted to trials at low overall risk of bias, we performed exploratory meta-epidemiologic analyses to compare treatment outcomes between trials at high or unclear risk of bias and trials at low risk of bias for each key domain of the risk of bias tool and for the overall risk of bias using the same methodology as described above. An ROR greater than 1 indicates larger treatment outcomes for trials at high or unclear risk of bias than trials at low risk of bias.

We used Stata SE version 11.0 (StataCorp) and R version 3.0.2 (R Foundation for Statistical Computing [http://www.R-project.org]) for statistical analysis. P < .05 (2-sided) was set as the level of significance.

Results

General Characteristics of the Meta-analyses

Of the 163 meta-analyses (1240 RCTs), 92 (705 RCTs) assessed a subjective outcome and 71 (535 RCTs) an objective one. The characteristics of each meta-analysis are reported in eTable 3 in the Supplement for meta-analyses of subjective outcomes and eTable 4 in the Supplement for meta-analyses of objective outcomes; their references are in the eReference list in the Supplement. Briefly, the median number of contributing trials was 6 (range, 3-48) for meta-analyses of subjective outcomes and 6 (range, 3-25) for those of objective outcomes. With all available trials included, we found a statistically significant association between the experimental treatment and outcomes in 60 of 92 (65%) meta-analyses of subjective outcomes and 24 of 71 (34%) meta-analyses of objective outcomes (Table 1).

Comparison of Treatment Outcomes Between Meta-analysis of All Trials and Alternative Strategies for Analysis

Meta-analysis of All Trials vs Single Most Precise Trial

Treatment outcomes were, on average, larger for the meta-analysis of all trials than for the single most precise trial, with a combined ROR of 1.13 (95% CI, 1.07-1.19, P < .001) for subjective outcomes and 1.03 (95% CI, 1.01-1.05, P = .002) for ob-
Objective outcomes. Heterogeneity across meta-analyses was low for both analyses ($P = 0\%$) (eFigure 2 in the Supplement and Table 2).

The difference in treatment outcomes between these 2 strategies was deemed substantial for 47 of 92 (51\%) meta-analyses of subjective outcomes (meta-analysis of all trials showing larger outcomes in 40/47) and 28 of 71 (39\%) meta-analyses of objective outcomes (meta-analysis of all trials showing larger outcomes in 21/28). For example, in a meta-analysis assessing the association between direct stenting and a composite of death or myocardial infarction, the ROR was 1.78 (95\% CI, 1.04-3.07), with an OR of 0.77 (95\% CI, 0.60-0.97) for the meta-analysis of all trials and 1.37 (95\% CI, 0.75-2.47) for the single most precise trial.22

Meta-analysis of All Trials vs Meta-analysis Restricted to the Largest Trials

When comparing meta-analysis of all trials with meta-analysis of the largest trials, the ROR was 1.08 (95\% CI, 1.04-1.13, $P < .001$) for subjective outcomes and 1.03 (95\% CI, 1.00-1.06, $P = .044$) for objective outcomes. Heterogeneity across meta-analyses was moderate with subjective outcomes and low

| Table 1. Characteristics of the 163 Meta-analyses by Type of Outcome (Subjective vs Objective) |
|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|
| Characteristics of Meta-analyses               | Subjective (n = 92 [705 RCTs]) | Objective (n = 71 [535 RCTs]) |
| Cochrane review, No. (%)                        | 68 (74)                           | 50 (70)                           |
| Year of publication, No. (%)                    |                                  |                                  |
| 2008-2009                                        | 10 (11)                           | 14 (20)                           |
| 2010-2011                                        | 40 (43)                           | 29 (41)                           |
| 2012-2013                                        | 42 (46)                           | 28 (39)                           |
| No. of contributing trials                      |                                  |                                  |
| Median (range)                                   | 6 (3-48)                          | 6 (3-25)                          |
| ≥10 trials, No. (%)                              | 16 (17)                           | 15 (21)                           |
| Treatment outcome with the meta-analysis of all trials |                                  |                                  |
| OR (95\% CI), range                              | 0.05 (0.01-0.20) to 1.59 (1.15-2.20) | 0.14 (0.05-0.39) to 1.16 (0.63-2.15) |
| Statistical difference, No. (%)                 |                                  |                                  |
| In favor of experimental group                   | 60 (65)                           | 24 (34)                           |
| In favor of control group                        | 1 (1)                             | 0                                 |
| None                                             | 31 (34)                           | 47 (66)                           |
| Sample size of included trials                   |                                  |                                  |
| Minimum size per meta-analysis, median (range)   | 42 (13-512)                       | 41 (9-504)                        |
| Maximum size per meta-analysis, median (range)   | 295 (34-564 642)                  | 400 (62-48 840)                   |
| Meta-analysis with ≥1 trial with sample size >1000 patients, No. (%) | 18 (20)                           | 22 (31)                           |
| Overall risk of bias of included trials          |                                  |                                  |
| No. of trials at low risk per meta-analysis, median (range) | 0 (0-12)                         | 1 (0-9)                           |
| Meta-analysis with ≥1 trial at low risk of bias, No. (%) | 41 (45)                           | 40 (56)                           |

| Table 2. Summary of the Average Differences in Treatment Outcomes Between the Meta-analysis of All Trials and Each Alternative Strategy, Expressed as Ratios of Odds Ratios, by Type of Outcome (Subjective vs Objective) |
|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|
| Alternative Strategy                             | Subjective (n = 92 [705 RCTs]) | Objective (n = 71 [535 RCTs]) |
| ROR (95\% CI)*                                   | P Value | P (%)   | ROR (95\% CI)*                                   | P Value | P (%)   |
| Single most precise trial                        | 1.13 (1.07-1.19)                      | <.001   | 0.03 (1.01-1.05)                                  | .002    | 0       |
| Meta-analysis restricted to the largest trialsb  | 1.08 (1.04-1.13)                      | <.001   | 1.03 (1.00-1.06)                                  | .044    | 0       |
| Limit meta-analysis                               | 1.17 (1.11-1.22)                      | <.001   | 1.13 (0.82-1.55)                                  | .46     | 96      |
| Meta-analysis restricted to trials at low overall risk of bias | 0.94 (0.86-1.04) | .23     | 1.03 (1.00-1.06)                                  | .048    | 23      |

Abbreviations: OR, odds ratio; RCT, randomized clinical trial.

*An ROR greater than 1 indicates larger treatment outcomes with the meta-analysis of all trials than with the alternative strategy.

bThe largest trials are defined as those in quarter 4 of sample size within each meta-analysis.
with objective outcomes ($I^2 = 27\%$ and $0\%$, respectively) (eFigure 3 in the Supplement and Table 2). The difference in treatment outcomes between these 2 strategies was deemed substantial for $38$ of $92$ ($41\%$) meta-analyses of subjective outcomes (meta-analysis of all trials showing larger outcomes in 23/38) and 19 of 71 ($27\%$) meta-analyses of objective outcomes (meta-analysis of all trials showing larger outcomes in 8/19). For example, in a meta-analysis assessing the association between angiotensin-converting enzyme inhibitor used as secondary prevention after cardiovascular and recurrence of atrial fibrillation, the ROR was $1.80$ ($95\%$ CI, $1.17$-$2.78$), with an OR of $0.55$ ($95\%$ CI, $0.35$-$0.87$) for the meta-analysis of all trials and $0.99$ ($95\%$ CI, $0.81$-$1.20$) when restricting to the largest trials.23

**Meta-analysis of All Trials vs Limit Meta-analysis**

When comparing meta-analysis of all trials with limit meta-analysis, the combined ROR was $1.17$ ($95\%$ CI, $1.11$-$1.22$, $P < .001$) for subjective outcomes and $1.13$ ($95\%$ CI, $0.82$-$1.55$, $P = .46$) for objective outcomes. Heterogeneity across meta-analyses was low for subjective outcomes ($I^2 = 0\%$) and considerable for objective outcomes owing to 1 meta-analysis outlier ($I^2 = 96\%$) (eFigure 4 in the Supplement and Table 2). The exclusion of this outlier yielded an ROR of $1.13$ ($95\%$ CI, $1.08$-$1.19$, $P < .001$) with no detectable heterogeneity ($I^2 = 0\%$).

A sensitivity analysis based on meta-analyses including 10 trials or more yielded an ROR of $1.24$ ($95\%$ CI, $1.12$-$1.36$, $P < .001$) for subjective outcomes and $1.10$ ($95\%$ CI, $1.01$-$1.19$, $P = .02$) for objective outcomes (eFigure 5 in the Supplement).

The difference in treatment outcomes between these 2 strategies was deemed substantial for 62 of 92 ($67\%$) meta-analyses of subjective outcomes (meta-analysis of all trials showing larger outcomes in 51/62) and 39 of 71 ($55\%$) meta-analyses of objective outcomes (meta-analysis of all trials showing larger outcomes in 28/39). For example, in a meta-analysis assessing the association between psychological interventions and depression, the ROR was $1.54$ ($95\%$ CI, $1.23$-$1.94$), with an OR of $0.74$ ($95\%$ CI, $0.59$-$0.93$) for the meta-analysis of all trials and $1.14$ ($95\%$ CI, $0.84$-$1.55$) for the limit meta-analysis.24

**Effect of Alternative Strategies on Statistical Significance Observed in Meta-analysis of All Trials**

As supplementary data, we also assessed how often the alternative strategies eliminated the statistical significance observed with the meta-analysis of all trials and how often the alternative strategies turned a nonstatistically significant result into a statistically significant one. These results are presented in eFigure 7 in the Supplement.

**Comparison of Treatment Outcomes Between Trials at High or Unclear Risk of Bias and Those at Low Risk of Bias**

When exploring the different domains of the risk of bias tool, treatment outcomes were larger for trials at high or unclear risk of bias than for those at low risk for the domains sequence generation, allocation concealment, and blinding for both subjective and objective outcomes. We did not find any evidence of difference in treatment outcomes between trials at high or unclear overall risk and trials at low overall risk of bias within meta-analyses ($ROR, 0.96$ [95% CI, $0.87$-$1.08$] for subjective outcomes and $0.97$ [95% CI, $0.86$-$1.10$] for objective outcomes) (Figure).

**Discussion**

In this study, we compared estimated treatment outcomes between meta-analysis of all trials, the most common strategy, and alternative strategies based on trial size and on risk of bias. Treatment outcome estimates differed depending on the analytic strategy used, with treatment outcomes frequently being larger with meta-analysis of all trials than with the single most precise trial, meta-analysis of the largest trials, and limit meta-analysis. This finding seems to be more marked for subjective than objective outcomes. In contrast, we did not find any difference in treatment outcomes by overall risk of bias.

Systematic reviews of RCTs are considered by some to be the gold standard for assessing the efficacy of an intervention.2,26 Within systematic reviews, meta-analyses are extremely important as a way to summarize information into a single estimate.27 However, combining data in a meta-analysis results in a conflict between 2 principles: first, to include all available evidence, and second, to get the “best” estimate.28 In this study, we compared meta-analysis of all trials with several “best-evidence” alternative strategies and found that estimated treatment outcomes differed depending on the strategy used. We cannot say which strategy is the best because, as outlined by Ioannidis,29 we cannot know with 100% certainty the truth in any research question. Nevertheless, our results raise important questions about meta-analyses and outline the need to rethink certain principles.
In the 1990s, there was important debate on the ability of meta-analyses to predict the “true” treatment outcome. Some studies scrutinized discordances between meta-analyses and large randomized trials, the latter being considered the gold standard. Many authors warned against performing meta-analyses including mainly small-sized trials and recommended systematic sensitivity analyses to test the robustness of findings. Accumulating evidence concerning characteristics associated with treatment outcomes has supported these recommendations. Concerning trial size, several studies found that small and moderate-sized trials showed larger treatment outcomes as compared with the largest trials within meta-analyses. These larger treatment outcomes may be related to reporting bias (smaller trials being more prone to publication bias) or to outcome reporting bias but also to methodological differences between small and large trials or to inclusion of more homogeneous populations of patients in smaller trials.

Meta-epidemiologic studies have also yielded evidence that certain trial-level characteristics—allocation concealment, blinding, or exclusion of patients from analysis—are associated with overestimated treatment outcomes in meta-analyses. Despite this, reports seldom describe an evaluation of the robustness of results by sensitivity analysis based on risk of bias or an evaluation of small-study effect by funnel plots. Our results raise questions about the overall risk of bias, summarizing risk of bias across domains, as currently defined. The risk of bias tool includes methodological characteristics or domains shown to be associated individually with treatment outcomes in meta-epidemiologic studies. In contrast, no meta-epidemiologic study has assessed the effect of the overall risk of bias on treatment outcomes. In our study, treatment outcomes were larger for trials at high or unclear risk of bias than for trials at low risk for sequence generation, allocation concealment, and blinding, which is consistent with the BRANDO study combining data from several meta-epidemiologic studies. However, we did not find any differences in treatment outcomes by overall risk of bias. Despite being attractive, the use of an overall risk of bias combining the different domains is challenging. All domains may not have the same weight for risk of bias and may be associated with one another. Moreover, according to the current definition, trials with 1 domain at high risk and those with all key domains at high risk have the same risk of bias, whereas one may assume that the greater the number of domains at high risk of bias the greater the probability of biased results. The use of an overall score may also obscure differences related to specific aspects of study design or execution in specific settings. Jüni et al demonstrated years ago, in a study comparing the effects of various measures of quality, that weighting schemes used for quality scales were problematic. Further research is needed to explore whether one can obtain a simple measure of the overall risk of bias for a given trial and, if so, how.

Practical Recommendations

We recommend that authors of meta-analyses systematically assess the robustness of their results by performing sensitivity analyses. We suggest the comparison of the meta-analysis result to the result for the single most precise trial or meta-analysis of the largest trials and careful interpretation of the meta-analysis result if they disagree. If 10 trials or more are included, performing a limit meta-analysis as a sensitivity analysis would also be of interest.
We also recommend assessing the influence on treatment outcomes of each domain of the risk of bias tool separately rather than summarizing these domains into an overall risk of bias.

Limitations

Our sample of meta-analyses is not representative of all published meta-analyses. We used data from 3 collections of meta-analyses. The first collection (29% of the whole sample) included meta-analyses published in journals with the 10 highest impact factors for each medical specialty, for a more homogeneous sample. However, even when restricting our sample to the journals with the highest impact factor for each medical specialty, there could be a wide quality range. The 2 other collections were published in the Cochrane Database of Systematic Reviews. Some studies previously showed that Cochrane reviews were more likely to use more rigorous methods and have better reporting than non-Cochrane reviews.54,55 Moreover, our sample included meta-analyses published between 2008 and 2013, so it does not represent the most recent literature.

Conclusions

Our results show that estimating treatment outcomes in meta-analyses differs depending on the analysis strategy used. This instability in findings can result in major alterations in the conclusions derived from the analysis and underlines the need for systematic sensitivity analyses.

REFERENCES


43. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. J Clin Epidemiol. 2011;64(10):1060-1069.


