Draft Methods Report
Number XX

Empirical and simulation-based comparison of univariate and multivariate meta-analysis for binary outcomes

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD20850
www.ahrq.gov

This information does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services determination or policy.

Contract No.

Prepared by:
[Redacted]

Investigators
[Redacted]

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted, for which further reproduction is prohibited without the specific permission of copyright holders.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Authors [redacted] are involved in developing open source software for meta-analysis, but this software is available at no charge and the authors receive no financial benefit.

This report is based on research conducted under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD. The findings and conclusions in this document are those of the author(s), who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care researchers and funders of research make well-informed decisions in designing and funding research and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of scientific judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical research and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances.

This report may be used, in whole or in part, as the basis for research design or funding opportunity announcements. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

AHRQ Contacts

Elisabeth U. Kato, MD, MRP

Medical Officer

Center for Outcomes and Evidence

Agency for Healthcare Research and Quality

540 Gaither Road

Rockville, MD 20850

Contents

Abstract 9

Background 10

Methods 11

Aim 1. Formation of the database 12

Aim 2. Comparison between univariate and multivariate meta-analysis 13

Meta-analysis models 13

Model fitting 17

Recorded results 19

Aim 3. Simulation analysis 19

Simulation parameters 19

Generation of random data 20

Metrics 20

Results 24

Description of the database (aim 1) 24

Empirical comparisons (aim 2) 29

Normal approximation methods 29

Meta-analyses using the binomial or multinomial likelihood 39

Illustrative simulation analyses (aim 3) 45

Comparison between univariate and multivariate meta-analysis in simulations 48

Discussion 52

References 53

Tables

Table 1. Observed counts for two mutually exclusive outcomes (study k) 12

Table 2. Observed counts for two outcomes having an is-subset-of relationship (study k) 12

Table 3. True probabilities for experiencing two mutually exclusive outcomes (study k) 14

Table 4. True probabilities for the mutually exclusive categories implied by two outcomes that have an is-subset-of relationship (study k) 14

Table 5. True probabilities for two outcomes that have an is-subset-of relationship (study k) as a function of the mutually exclusive quantities in Table 4 14

Table 6. Simulation parameters 19

Table 7. Definition of MSE, bias and coverage probability 20

Table 8. Pseudo-algorithm for generating simulation data for mutually exclusive outcomes 22

Table 9. Pseudo-algorithm for generating simulation data for outcomes that have an is-subset-of relationship 23

Table 10. Thirty eight pairs of outcomes having an is-subset-of relationship 25

Table 11. Five pairs of mutually exclusive outcomes 28

Table 12. Two triplets of outcomes having an is-subset-of relationship 28

Table 13. Topics with RCTs with singular covariance matrices 29

Table 14. Estimated between-study correlation between the first and second outcome 38

Table 15. Posterior median and 95% credible interval for between-study correlations with uninformative priors and priors informative on the correlation sign 39

Table 16. Influential simulation parameters for random effects bivariate meta-analysis of two mutually exclusive outcomes 46

Table 17. Influential simulation parameters for random effects bivariate meta-analysis of two outcomes with an is-subset-of relationship 46

Table 18. MSE, coverage and bias for random effects bivariate meta-analysis (two mutually exclusive outcomes) 47

Table 19. MSE, coverage and bias for random effects bivariate meta-analysis (two outcomes that have an is-subset-of relationship) 47

Figures

Figure 1. Plot for choosing the regularizing constant for the ridge regression (example, topic 2) . 30

Figure 2. Summary estimates with univariate and bivariate meta-analysis in topics 1 through 43 (REML random effects, normal approximation) 31

Figure 3. Summary estimates with univariate and trivariate meta-analysis in topics 44 and 45 (REML random effects, normal approximation) 32

Figure 4. Relative summary odds ratios with univariate and bivariate meta-analyses in topics 1 through 43 33

Figure 5. Treatment effects with univariate and trivariate meta-analysis in topics 44 and 45. 34

Figure 6. Ratio of standard errors for log relative odds ratios from multivariate and univariate meta-analysis and estimated correlation coefficient between means 34

Figure 7. Difference in the point estimate for the first outcome with univariate and multivariate meta-analysis versus the variability in the elements of the within-study covariance matrix (topics 1 through 45) 36

Figure 8. Difference in the point estimate for the second outcome with univariate and multivariate meta-analysis versus the variability in the elements of the within-study covariance matrix (topics 1 through 45) 37

Figure 9. Difference in the point estimates with univariate and multivariate meta-analysis versus the number of studies reporting the other outcome (topics 1 through 45) 38

Figure 10. Comparison of univariate and multivariate meta-analysis using the binomial or the multinomial distribution to model within-study variance (topics 1 through 43) 41

Figure 11. Comparison of univariate and trivariate meta-analysis using the binomial or the multinomial distribution to model within-study variance (topics 44 and 45) 42

Figure 12. Relative summary odds ratios with univariate and bivariate meta-analyses using the binomial or the multinomial distribution to model within-study variance (topics 1 through 43) ... 43

Figure 13. Relative summary odds ratios with univariate and trivariate meta-analyses using the binomial or the multinomial distribution to model within-study variance (topics 44 and 45) 44

Figure 14. Summary odds ratios with univariate and bivariate meta-analysis across all simulations (mutually exclusive outcomes) 48

Figure 15. Summary odds ratios with univariate and bivariate meta-analysis across all simulations (outcomes with an is-subset-of relationship) 49

Figure 16. Standard errors of summary odds ratios with univariate and bivariate meta-analysis across all simulations (mutually exclusive outcomes) 49

Figure 17. Standard errors of summary odds ratios with univariate and bivariate meta-analysis across all simulations (outcomes with an is-subset-of relationship) 50

Figure 18. Comparison of standard errors of the difference in the log summary odds ratios of two mutually exclusive outcomes with univariate and bivariate meta-analysis 51

Figure 19. Comparison of standard errors of the difference in the log summary odds ratios of two outcomes that have an is-subset-of relationship with univariate and bivariate meta-analysis 51

Appendix

The Appendix includes information on regularized regression for dealing with singular covariance matrices, formulas for the covariances between odds ratios of correlated outcomes, and tables with additional results.

Appendices cited in this report are available at http://ahrq.gov/XXXXX

[to be completed by AHRQ]

Abstract

Background: Many questions in evidence-based medicine involve multiple outcomes. They can be approached with separate, independent meta-analyses, or they can be analyzed jointly, in a single model. We aimed to compare univariate with multivariate meta-analysis in real examples and in an illustrative simulation study.

Methods: We screened the whole Cochrane Library of Systematic Reviews (2012, first quarter) to identify sets of univariate meta-analyses of categorical outcomes that can also be analyzed jointly. Eligible were pairs or triplets of meta-analyses comparing the same interventions; having at least seven randomized controlled trials (RCTs) reporting all outcomes; and in which the numbers in the cross-classification of outcomes were exactly recoverable. Examples of the latter are outcomes that are mutually exclusive, or sets outcomes where the one is a subset of the other. We analyzed these data with univariate and multivariate meta-analysis. In an accompanying simulation study, we compared summary estimates and their standard errors with univariate and multivariate meta-analysis.

Results: We identified 45 pairs or triplets of binary meta-analyses corresponding to 1473 RCTs and 258,675 randomized patients. In 38 (of 45) topics the first outcome was a subset of the second outcome; in five pairs the outcomes were mutually exclusive, and in two triplets the outcomes had an is-subset-of relationship. The 45 topics pertained to various medical areas (e.g., cardiology, surgery, mental health). Overall, the summary effects for each outcome and the accompanying confidence/credible intervals were very similar with univariate and multivariate meta-analysis (using the approximate and discrete likelihood). The methods yield different confidence intervals when one calculates differences of meta-analysis summary treatment effects across the outcomes of interest. Depending on the estimated covariance between the compared effects, the multivariate methods can yield tighter or wider confidence intervals than univariate methods. Most likely, systematic review conclusions from the meta-analyses in the empirical sample would not change with the method of analysis. The simulation analyses were congruent with the aforementioned observations from the empirical analyses.

Conclusions: In the empirical sample and the simulation study, the numerical difference in the summary effects and their confidence intervals between univariate and multivariate meta-analysis was almost always small. It is likely that in many (if not most) cases conclusions would not change with the method of analysis. The choice between univariate and multivariate meta-analysis may have more academic than practical interest.

Background

The growing number of treatment choices, as well as the rapid escalation in the cost of health care, has spawned a need for scientifically rigorous comparisons of the efficacy and safety of drugs, devices and treatments in clinical practice. To date, most quantitative comparisons carried out by the Evidence Based Practice Centers (EPCs) funded by AHRQ have relied on traditional meta-analysis comparing two treatments with respect to a single outcome. However, many questions involve multiple outcomes. Standard assessments have approached these questions with separate meta-analyses, leading to a plethora of analyses to interpret without any quantitatively rigorous methods for integrating them. As these are multivariate problems, multivariate statistical methods offer a solution.[1,2]

Arguably, joint analysis of all relevant information may be preferable to separate analyses that use only a subset of the available information. For example, imagine that we have *K* studies comparing statins versus no statins for the outcomes of cardiac and noncardiac mortality. The usual approach is to perform two separate meta-analyses: one for cardiac mortality and one for noncardiac mortality. However, both outcomes are evaluated in the same patients (same studies), and are thus stochastically dependent (correlated). Intuitively, knowing something about one outcome tells us something about the other. In the previous example, if too many people die of cardiac causes, fewer people are at risk of dying of noncardiac causes: the proportions dying of the two mutually exclusive causes are negatively correlated. By analyzing the two outcomes jointly we can capitalize on these correlations – an opportunity that is lost with separate, univariate meta-analyses.[2,3]

Statistical methods for simultaneously analyzing multiple outcomes have appeared mainly in the past decade, along with applications to important problems. [2-9] Because they are new, these methods have not migrated into standard clinical research and remain the province of experts. Therefore the empirical evidence base for comparing multivariate meta-analysis of multiple outcomes or multiple follow-ups versus separate univariate meta-analyses is limited. Such empirical data are useful for making informed methodological recommendations regarding the use of multivariate versus univariate meta-analysis in applied systematic review and technology assessment. For example, if the numerical difference in the summary effects and their confidence intervals is always very small, the choice between univariate and multivariate meta-analysis has more academic than practical interest. If nonnegligible differences are common, recommendations can differ.

In this project we perform an empirical evaluation of separate (univariate) versus joint (multivariate) meta-analysis for comparing two treatments with respect to two or more categorical outcomes using real data from the Cochrane Library of Systematic Reviews. We supplement the empirical observations with a simulation study of a large number of scenarios representative of actual analyses.

Project aims

This project has three aims:

Aim 1: Assemble a database of examples from the Cochrane Library of Systematic Reviews that can be analyzed with multivariate meta-analyses using only information reported in the reviews. (Cochrane systematic reviews perform only univariate meta-analysis.)

Aim 2: Perform an empirical comparison of results with univariate versus multivariate meta-analysis for the database of Aim 1.

Aim 3: Perform illustrative simulation analyses to compare multivariate and separate univariate meta-analyses for the case of two categorical outcomes that have a mutually exclusive relationship, or an is-subset-of relationship.

Aim 1 provides data for the quantitative calculations of Aim 2 and informs the choice of parameters for the simulations of Aim 3.

Methods

Having information on within-study correlations is necessary for the joint meta-analysis of two or more outcomes. For some problems, such information is not extractable from data reported in published studies, and one has to obtain individual patient data (for example, see Peter et al.[10]) or to impute the missing information based on prior knowledge, as was done in the example in Berkey et al.[6]. (This is true for all joint meta-analyses of continuous outcomes and for many joint meta-analyses between categorical outcomes.) For other problems (all of which pertain to categorical outcomes) the numbers in the cross-classification of outcomes are exactly recoverable. Thus, one can calculate within-study correlations (or equivalent information) from typically reported data. These are cases where:

1. The outcomes are mutually exclusive, that is, a person can experience only one of them. Examples are the pair of outcomes “death from breast cancer” and “death from causes other than breast cancer” and the pair “births by caesarean section” and “spontaneous vaginal births”.[3]
2. The people experiencing one outcome are a subset of the people experiencing the other outcome. Such is the pair of outcomes “survival at 6 months” and “survival at 12 months”, because those alive at 12 months are a subset of those who were alive at 6 months. Similarly, “withdrawals due to adverse events” are a subset of “withdrawals for any reason”.[9]

Table 1 shows an example of mutually exclusive outcomes, and Table 2 an example of outcomes that have an is-subset-of relationship. The examples extend to more than two outcomes in the obvious way.

Table 1. Observed counts for two mutually exclusive outcomes (study k)

Arm	*Outcome 1: Breast cancer deaths*	*Outcome 2: Death from other causes*	*Alive (remaining)*	*Total*
Comparator (C)	X_{k1}	X_{k2}	X_{k3}	$N_k^{(C)} = X_{k1} + X_{k2} + X_{k3}$
Treatment (T)	Y_{k1}	Y_{k2}	Y_{k3}	$N_k^{(T)} = Y_{k1} + Y_{k2} + Y_{k3}$

Table 2. Observed counts for two outcomes having an is-subset-of relationship (study k)

Arm	*Outcome 1: Alive at 12 months*	*Outcome 2: Alive at 6 months*	*Total*
Comparator (C)	A_{k1}	$A_{k2}(\geq A_{k1})$	$N_k^{(C)}$
Treatment (T)	B_{k1}	$B_{k2}(\geq B_{k1})$	$N_k^{(T)}$

The outcomes "alive at 12 months" and "alive at 6 months" imply the following three mutually exclusive categories: "alive at 12 months", "dying between 6 and 12 months", and "dying between 0 and 6 months". Denote the counts for the latter three categories in the comparator by F_{k1}, F_{k2} , and F_{k3} and in the treatment by G_{k1}, G_{k2}, and G_{k3} respectively. Then $F_{k1} = A_{k1}$, $F_{k1} + F_{k2} = A_{k2}$, and $F_{k1} + F_{k2} + F_{k3} = N_k^{(C)}$, and analogously for the treatment arm.

Aim 1. Formation of the database

Eligibility criteria

A single investigator [##name redacted##] screened the Cochrane Database of Systematic Reviews, Quarter 1, 2012, to identify pairs or triplets of univariate meta-analyses for which numbers on cross-classification of outcomes are exactly recoverable, and that can be meta-analyzed jointly. Eligible were pairs or triplets of univariate meta-analyses

1. comparing the same interventions;
2. having at least seven randomized controlled trials, RCTs,[a] (or at least half of the RCTs if the number of studies *K*>7) with at least two events and at least 10 patients per arm;
3. where an adequate number of RCTs reported all outcomes of interest.[b] This was defined as at least seven RCTs with two or more events per arm (or at least half of the RCTs if *K*>7); and
4. reporting actual data rather than sensitivity analyses (such as "worst case scenario" analyses).

We excluded Cochrane reviews that have been withdrawn. We did not consider collections of studies that have not been pooled quantitatively by the primary Cochrane reviewers

[a] Although the vast majority of studies in Cochrane reviews are randomized, a few are not. Nevertheless, we refer to all studies in Cochrane meta-analyses as "RCTs".
[b] Paragraph "On the number of parameters to be estimated" (page 17) explains the rationale for choosing *K=7* as the minimum number of studies in a meta-analysis.

(irrespective of rationale). When the primary Cochrane reviewers performed quantitative syntheses within subgroups of studies but not across these subgroups, we considered each subgroup as a distinct univariate meta-analysis. When the primary reviewers pooled across subgroups, we ignored subgroup classifications and considered the overall synthesis as a single meta-analysis. We excluded meta-analyses of survival outcomes based on approximating a log hazard ratio using the numbers of observed minus expected events.

Some Cochrane reviews could contribute more than one otherwise eligible pair or triplet of outcomes. When applicable, we preferred to form triplets of meta-analyses to selecting one of the three possible pairs in a triplet. To avoid duplication of information, we included only independent pairs or triplets of univariate meta-analyses from a single Cochrane review. These were defined as pairs or triplets pertaining to different intervention comparisons, or pairs or triplets that included nonoverlapping collections of RCTs.

Among nonindependent pairs or triplets of outcomes, we selected the one with the largest number of RCTs reporting all outcomes. We broke any ties by selecting the pair or triplet of meta-analyses with the largest total number of randomized patients, then the largest total number of events, and then randomly.

Data extraction

For each eligible pair or triplet of univariate meta-analyses we recorded the title and identification code of the parent Cochrane review, the compared interventions, outcomes, subgroup definitions (if applicable), the Cochrane Library identification numbers of the included RCTs, and the first author and the year of publication of each RCT. We also recorded the number of events for each outcome and each arm. Finally, we categorized the topics of eligible reviews as pertinent to general medicine, pediatrics, obstetrics and gynecology, nephrology, mental health, cardiology, infectious diseases, and surgery.

Aim 2. Comparison between univariate and multivariate meta-analysis

For each eligible pair or triplet of meta-analyses of outcomes we performed univariate (separate) and multivariate (joint) meta-analyses with both fixed and random effects. We used two modeling approaches: the first modeled within-study variability with normal distributions, and the second used discrete distributions (binomial for univariate meta-analysis, multinomial for multivariate). Because most meta-analyses involve clinical and methodological diversity, we report results from random effects analyses in the main text of the report. We report results from fixed effects analyses in the Appendix.

Meta-analysis models

Assume that we have K studies (indexed by k) reporting M outcomes (indexed by m). The individual outcomes are either mutually exclusive and associated with a single period of follow-up, or have an is-subset-of relationship. Continuing with the example in Table 1, Table 3 shows the true (population) probabilities for the categories defined by outcomes that are mutually exclusive.

Table 3. True probabilities for experiencing two mutually exclusive outcomes (study k)

Arm	*Outcome 1: Breast cancer deaths*	*Outcome 2: Death from other causes*	*Alive (remaining)*
Comparator (C)	$\pi_{k1}^{(C)}$	$\pi_{k2}^{(C)}$	$\pi_{k3}^{(C)} = 1 - \pi_{k1}^{(C)} - \pi_{k2}^{(C)}$
Treatment (T)	$\pi_{k1}^{(T)}$	$\pi_{k2}^{(T)}$	$\pi_{k3}^{(T)} = 1 - \pi_{k1}^{(T)} - \pi_{k2}^{(T)}$

This table corresponds to Table 1 (observed counts).

Sets of outcomes that have an is-subset-of relationship imply a set of mutually exclusive categories (Table 4). The true probabilities for experiencing outcomes that have an is-subset-of relationship are summations over the cells of Table 4, as defined in Table 5. One can extend the notation to more than two outcomes in the obvious way.

Table 4. True probabilities for the mutually exclusive categories implied by two outcomes that have an is-subset-of relationship (study k)

Arm	*Alive at 12 months*	*Dying between 6 and 12 months*	*Dying between 0 and 6 months*
Comparator (C)	$\gamma_{k1}^{(C)}$	$\gamma_{k2}^{(C)}$	$\gamma_{k3}^{(C)} = 1 - \gamma_{k1}^{(C)} - \gamma_{k2}^{(C)}$
Treatment (T)	$\gamma_{k1}^{(T)}$	$\gamma_{k2}^{(T)}$	$\gamma_{k3}^{(T)} = 1 - \gamma_{k1}^{(T)} - \gamma_{k2}^{(T)}$

Table 5. True probabilities for two outcomes that have an is-subset-of relationship (study k) as a function of the mutually exclusive quantities in Table 4

Arm	*Outcome 1: Alive at 12 months*	*Outcome 2: Alive at 6 months*
Comparator (C)	$\lambda_{k1}^{(C)} = \gamma_{k1}^{(C)}$	$\lambda_{k2}^{(C)} = \gamma_{k1}^{(C)} + \gamma_{k2}^{(C)}$
Treatment (T)	$\lambda_{k1}^{(T)} = \gamma_{k1}^{(T)}$	$\lambda_{k2}^{(T)} = \gamma_{k1}^{(T)} + \gamma_{k2}^{(T)}$

The γ's are mutually exclusive and exhaustive within each treatment arm (Table 4), but the λ's are not. This table corresponds to Table 2 (observed counts).

Univariate fixed-effect meta-analysis

Structural model

For mutually exclusive outcomes, the true effect size for outcome m (here, the log odds ratio, log OR) in study *k* is

$$\theta_{km} = \text{logit}(\pi_{km}^{(T)}) - \text{logit}(\pi_{km}^{(C)}) \tag{1}$$

For outcomes that have an is-subset-of relationship, write

$$\theta_{km} = \text{logit}(\lambda_{km}^{(T)}) - \text{logit}(\lambda_{km}^{(C)}) \tag{2}$$

Under a fixed effects meta-analysis model, the true effect for each outcome is the same in all K studies

$$\theta_{km} = \theta_{m.} \tag{3}$$

Observational model

One has two options in modeling within-study variation. The first option is to assume that the sample estimate of the log odds ratio, $\hat{\theta}_{km}$, is normally distributed

$$\hat{\theta}_{km} \sim \mathcal{N}(\theta_{m}, \sigma_{km}^{2}), \tag{4}$$

with the sample conditional variance σ_{km}^{2} assumed known and calculated from the data. (This assumption is often made in meta-analysis, but is not always justified.) For mutually exclusive outcomes

$$\hat{\theta}_{km} = \text{logit}\left(\frac{Y_{km}}{N_{km}^{(T)}}\right) - \text{logit}\left(\frac{X_{km}}{N_{km}^{(C)}}\right), \text{ and} \tag{5}$$

$$\sigma_{km}^{2} = \frac{1}{Y_{km}} + \frac{1}{N_{km}^{(T)} - Y_{km}} + \frac{1}{X_{km}} + \frac{1}{N_{km}^{(C)} - X_{km}}. \tag{6}$$

For outcomes with an is-subset-of relationship, substitute A_{km} for X_{km} and B_{km} for Y_{km} in (5) and (6).

Alternatively, one can model the number of events in Table 1 and Table 2 with binomial distributions. For the case of mutually exclusive outcomes

$$X_{km} \sim \mathcal{B}(N_{k}^{(C)}, \pi_{km}^{(C)}), \text{ and} \tag{7}$$

$$Y_{km} \sim \mathcal{B}(N_{k}^{(T)}, \pi_{km}^{(T)}). \tag{8}$$

For outcomes with an is-subset-of relationship

$$A_{km} \sim \mathcal{B}(N_{k}^{(C)}, \lambda_{km}^{(C)}), \text{ and} \tag{9}$$

$$B_{km} \sim \mathcal{B}(N_k^{(T)}, \lambda_{km}^{(T)}). \tag{10}$$

Univariate random-effects meta-analysis

For random effects we assume that the true effect in each study is normally distributed around a mean θ_m with a between-study variance τ_m^2. Thus (3) is replaced by

$$\theta_{km} \sim \mathcal{N}(\theta_m, \tau_m^2). \tag{11}$$

All other model-defining equations remain the same.

Multivariate fixed-effects meta-analysis

We are interested in the same comparisons as in the univariate case. As explained elsewhere, outcomes that are mutually exclusive or have an is-subset-of relationship are correlated and can be analyzed jointly.[3,9] The key is to capitalize on the multinomial structure of the data, which is outlined in Table 3 and in Table 4, respectively.

Structural model

Arrange the *M* true effects in study k in a column vector $\Theta_k = (\theta_{k1}, \ldots, \theta_{kM})'$, with prime (') denoting transpose and boldface denoting vectors (or matrices). The elements θ_{km} are defined by (1) for mutually exclusive outcomes and (2) for outcomes with an is-subset-of relationship. Under fixed effects, the true effects in all K studies are the same:

$$\Theta_k = \Theta. \tag{12}$$

Observational model

Write $\hat{\Theta}_k = (\hat{\theta}_{k1}, \ldots, \hat{\theta}_{kM})'$ for the estimate of Θ_k. In analogy to the univariate meta-analysis case, one can model within-study variability with a multivariate normal distribution

$$\hat{\Theta}_k \sim \mathcal{N}(\Theta_k, \Sigma_k), \tag{13}$$

where Σ_k is an $M \times M$ covariance matrix encoding the correlations between the outcomes within each study. The elements of Σ_k are assumed known and are calculated from the data using the formulas in the Appendix. (As noted in the univariate case, this assumption is not always justified.) Note that the formulas for the covariances are different for outcomes that are mutually exclusive compared with outcomes that have an is-subset-of relationship. The analogy between (3) and (12) and between (4) and (13) is obvious.

Alternatively, one can use the multinomial distribution to model observed counts in each arm. For mutually exclusive outcomes, we write using the notation in Table 1 and Table 3:

$$(X_{k1},\ldots,X_{kM}) \sim \mathcal{M}\left(N_k^{(C)},(\pi_{k1}^{(C)},\ldots,\pi_{kM}^{(C)})\right) \tag{14}$$

for the comparator arm, and

$$(Y_{k1},\ldots,Y_{kM}) \sim \mathcal{M}\left(N_k^{(T)},(\pi_{k1}^{(T)},\ldots,\pi_{kM}^{(T)})\right) \tag{15}$$

for the treatment arm. For outcomes that have an is-subset-of relationship, we have to model the counts for the mutually exclusive categories implied by the outcomes of interest. As an example, the outcomes "alive at 6 months" and "alive at 12 months" imply the following three mutually exclusive categories: "alive at 12 months", "dying between 6 and 12 months", and "dying between 0 and 6 months". Using the notation in Table 2 and Table 4, we write for the comparator and treatment arms

$$(F_{k1},F_{k2},F_{k3}) = (A_{k1},A_{k2}-A_{k1},N_k^{(C)}-A_{k2}) \sim \mathcal{M}\left(N_k^{(C)},(\gamma_{k1}^{(C)},\gamma_{k2}^{(C)},\gamma_{k3}^{(C)})\right), \text{ and} \tag{16}$$

$$(G_{k1},G_{k2},G_{k3}) = (B_{k1},B_{k2}-B_{k1},N_k^{(T)}-B_{k2}) \sim \mathcal{M}\left(N_k^{(T)},(\gamma_{k1}^{(T)},\gamma_{k2}^{(T)},\gamma_{k3}^{(T)})\right), \tag{17}$$

respectively. It is easy to extend (16) and (17) to more than two outcomes with an is-subset-of relationship.

Multivariate random-effects meta-analysis

In the random effects case, in place of (12) the vector of the true treatment effects Θ_k is assumed to follow a multivariate normal distribution

$$\Theta_k \sim \mathcal{N}(\Theta,\mathbf{T}), \tag{18}$$

where $\mathbf{T}$ is an $M \times M$ between-study covariance matrix. Trikalinos and Olkin 2008 and 2012 discuss parameterizations for $\mathbf{T}$.[3,9] In this work we use an unstructured $\mathbf{T} = (\tau_{\bar{y}})$.

Model fitting

Meta-analysis models using normal approximations

Univariate and multivariate meta-analyses using the normal distribution to model within-study variability were fitted with maximum likelihood for fixed effects, and restricted maximum likelihood for random effects. For details see Trikalinos and Olkin 2008 and 2012.[3,9] For studies with singular covariance matrices we used the ridge regression approach outlined in the appendix of Trikalinos and Olkin 2012.[9] The ridge regression is a form of regularized regression (details in

the Appendix). In sensitivity analyses we excluded studies with singular covariance matrices; the results were very similar or identical to those from the ridge regression analyses, and are not shown.

Meta-analysis models using the discrete likelihood

Univariate meta-analyses that use the binomial distribution to model counts in each study arm can be fit in the generalized linear mixed models framework using routines readily available in general statistical packages such as *xtmelogit* in Stata or *lmer* in R. However, the multivariate versions of the meta-analysis models specified here cannot be fit in these general routines. The available generalized linear mixed model (GLMM) packages in R, Stata and SAS do not allow the user to specify the random effects distribution in (18), where the random effects pertain to the log ORs. Instead they only allow specification of models in which the random effects are on the logit-transformed probabilities in Table 3 or Table 4.

Optimizing the likelihood for the random effects multinomial model outside a GLMM package is nontrivial, because it involves calculating complicated integrals numerically. Thus we did not develop routines for fitting this model. Instead we fitted the model using Markov Chain Monte Carlo (MCMC) methods in the Bayesian framework. For the Bayesian analyses we used vague (noninformative) prior distributions, as described in the Appendix, as well as prior distributions that were informative on the sign of the between-study correlation coefficients.

On the number of parameters to be estimated

For each univariate meta-analysis the parameters to be estimated are the true mean θ_m for fixed effects, and the true mean and the between-study variance τ_m^2 for random effects. For multivariate meta-analyses, the parameters to be estimated are the M true means (the elements of Θ) for fixed effects, and the $M(M+1)/2$ elements of the unstructured between-study covariance matrix $\mathbf{T}$. To fit the models, we must have more independent data points than parameters. From each study k reporting all outcomes of interest we have $M-1$ data points. So for random effects models, we must have $(M-1)K > M + M(M+1)/2$, or $K > M/(M-1) + M(M+1)/(2(M-1))$. For example, for $M = 2$, $K > 5$. The minimum K is probably not sufficient, and thus we opted to set as an eligibility criterion that meta-analyses should have at least seven studies reporting on all outcomes.

Software

(Restricted) maximum likelihood meta-analyses were performed in Stata using custom code (see the code in Trikalinos and Olkin 2012).[9] Bayesian meta-analyses were performed in JAGS (specifically, using the *rjags* library in R). We used three MCMC chains and a burn-in between 10000 and 50000 iterations. We monitored convergence with the Gelman-Rubin diagnostic for stochastic nodes corresponding to the meta-analysis means and the elements of their between-study covariance matrices. We declared convergence when the 97.5 percentile of the diagnostic was 1.10 or less for all monitored stochastic nodes, and provided that on visual inspection the traceplots of the MCMC chains were suggestive of good mixing.

Recorded results

From each analysis we recorded the summary effects for each outcome along with the respective variances (and covariances, if applicable), and results on between-study variances (covariances and correlations, as applicable). We also calculated pairwise differences between outcome effects, and the confidence intervals for these differences. These differences are the log relative odds ratios for observing one outcome versus the other, and are not necessarily helpful or informative for all contexts or topics; however, we calculated them for all topics, because this work makes technical rather than clinical conclusions and observations. We used the *t* distribution with *K-1* degrees of freedom to construct confidence intervals for estimated means and differences of estimated means.

For Bayesian analyses we recorded the median and the 0.025 and 0.975 quantiles of the posterior distribution for each quantity of interest. The latter two can be thought as being the endpoints of a 95% credible interval (95% CrI), the Bayesian analogue of a 95% confidence interval.

Aim 3. Simulation analysis

We generated illustrative simulation data as described in the next two paragraphs.

Simulation parameters

Table 6 shows the simulation parameters. These span a representative range of scenarios. Zero between-study standard deviations correspond to fixed-effects realities, and are probably not the norm in real-life applications. However, they are examined for completeness.

Table 6. Simulation parameters

#	*Parameter*	*Values*
1	Number of studies per meta-analysis, K_j	10, 20
2	Number of patients per study arm, $N^t_{j,k} = N^c_{j,k} = N_j$	50, 100, 500 [Same for all k, and for both study arms]
3	Vector of probabilities for the 3 categorical outcomes in the controls: • For mutually exclusive outcomes: $\pi^{(C)}_{j,k} = \pi^{(C)}_j$, or $(\pi^{(C)}_{j,k1}, \pi^{(C)}_{j,k2}, \pi^{(C)}_{j,k3}) = (\pi^{(C)}_{j,1}, \pi^{(C)}_{j,2}, \pi^{(C)}_{j,3})$ • For outcomes with an is-subset-of relationship: $\gamma^{(C)}_{j,k} = \gamma^{(C)}_j$, or $(\gamma^{(C)}_{j,k1}, \gamma^{(C)}_{j,k2}, \gamma^{(C)}_{j,k3}) = (\gamma^{(C)}_{j,1}, \gamma^{(C)}_{j,2}, \gamma^{(C)}_{j,3})$	(1/6, 1/6, 2/3) [Same for all k]
4	Marginal mean odds ratio for the first outcome, $\mu_{j,1}$	1, 1.5
5	Marginal mean odds ratio for the second outcome, $\mu_{j,2}$	1, 1.5

6	Between-study standard deviation for the log odds ratio of the first outcome, $\tau_{j,1}$	0, 0.1, 0.5
7	Between-study standard deviation for the log odds ratio of the second outcome, $\tau_{j,2}$	0, 0.1, 0.5
8	Between-study correlation between log-odds ratios for the two outcomes, ρ_j	0.25, 0.50, 0.75

j indexes simulation scenarios. Write $\boldsymbol{\mu}_j = (\mu_{j,1}, \mu_{j,2})'$, $\boldsymbol{\tau}_j = (\tau_{j,1}, \tau_{j,2})'$, and $\mathbf{R}_j = \begin{bmatrix} 1 & \rho_j \\ & 1 \end{bmatrix}$.

Generation of random data

For each scenario we generated 500 random sets of studies on two correlated binary outcomes. The following pseudo-algorithm describes the process for scenario *j* (all subscripts for the scenario have been dropped for notational simplicity). The pseudo-algorithms in Table 7 (for mutually exclusive outcomes) and Table 8 (for outcomes that have an is-subset-of relationship) explain the generation of simulation data.

We verified the fidelity of the simulations by comparing the means and covariance matrices of the empirical distributions versus the simulation parameters for the study effects $(\theta_{j,k1}, \theta_{j,k2})'$ and the proportions of events in the control arms.

Metrics

The main aim of the simulation study is to illustrate comparisons of results from univariate and multivariate meta-analysis. Nevertheless, we calculated the mean squared error (MSE), the bias and the coverage probability for all examined methods. We report these metrics for each outcome for analyses using the bivariate random effects method. Additional results are available from the authors.

Table 7. Definition of MSE, bias and coverage probability

Metric	*Formula*	*Description*	*Comment*
Mean squared error (MSE)	$\frac{1}{500}\sum_{i=1}^{500}(\hat{\mu}_{j,1i} - \mu_{j,1})^2$ $\frac{1}{500}\sum_{i=1}^{500}(\hat{\mu}_{j,2i} - \mu_{j,2})^2$	The average squared difference between the true (simulated) mean and its estimate across the 500 simulation replicates in scenario *j*.	• Desirable to have MSE near zero. • MSE can be high even if bias is 0, because positive and negative deviations of the estimates from the true mean do not cancel out. • MSE is the sum of the variance of an estimate plus the square of its bias.

Metric	*Formula*	*Description*	*Comment*
Bias	$\frac{1}{500}\sum_{i=1}^{500}(\hat{\mu}_{j,1i}-\mu_{j,1})$ $\frac{1}{500}\sum_{i=1}^{500}(\hat{\mu}_{j,2i}-\mu_{j,2})$	The average difference between the true (simulated) mean and its estimate across the 500 simulation replicates in scenario *j*.	• Desirable to have bias near zero.
Coverage probability	$\frac{1}{500}\sum_{i=1}^{500}I(\mu_{j,1}\in[95\%\text{CI of }\hat{\mu}_{j,1i}])$ $\frac{1}{500}\sum_{i=1}^{500}I(\mu_{j,2}\in[95\%\text{CI of }\hat{\mu}_{j,2i}])$	The proportion of times the 95% confidence interval of the estimated summary mean contains the true value.	• Desirable to have coverage near 95%. • Coverage higher than 95% indicates an inefficient estimator • Coverage less than 95% indicates an inaccurate estimator

$\hat{\mu}_{j,1i}$ stands for the meta-analysis estimate of $\mu_{j,1}$ and analogously for the other outcome.

Table 8. Pseudo-algorithm for generating simulation data for mutually exclusive outcomes

Choose a combination of values for the parameters in Table 6.

Do 500 times (for notational simplicity we drop the index for the simulation draw):

Draw *K* vectors of counts for the three categories of events in the comparator arms:

$$\mathbf{X}_{j,k} = (X_{j,k1}, X_{j,k2}, X_{j,k3}) \sim \mathcal{M}\left(N_j^{(C)}, (\pi_{j,k1}^{(C)}, \pi_{j,k2}^{(C)}, \pi_{j,k3}^{(C)})\right)$$

Draw *K* vectors of study effects for outcomes 1 and 2

$(\theta_{j,k1}, \theta_{j,k2})' = \boldsymbol{\theta}_{j,k} \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{T}\right)$, where $\mathbf{T} = \mathrm{diag}(\boldsymbol{\tau})\,\mathbf{R}\,\mathrm{diag}(\boldsymbol{\tau})$

Calculate the probabilities for the three categories of events in the treatment arms:

$$\pi_{j,k1}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k1} + \mathrm{logit}(\pi_{j,k1}^{(C)})\right),$$

$$\pi_{j,k2}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k2} + \mathrm{logit}(\pi_{j,k2}^{(T)})\right), \text{ and}$$

$$\pi_{j,k3}^{(T)} = 1 - \pi_{j,k1}^{(T)} - \pi_{j,k2}^{(T)}.$$

If $\pi_{j,k3}^{(T)} \leq 0$, set

$$\pi_{j,k3}^{(T)} := \frac{10^{-4}}{\pi_{j,k1}^{(T)} + \pi_{j,k2}^{(T)} + 10^{-4}},$$

$$\pi_{j,k2}^{(T)} := \frac{\pi_{j,k2}^{(T)}}{\pi_{j,k1}^{(T)} + \pi_{j,k2}^{(T)} + 10^{-4}}, \text{ and}$$

$$\pi_{j,k1}^{(T)} := \frac{\pi_{j,k1}^{(T)}}{\pi_{j,k1}^{(T)} + \pi_{j,k2}^{(T)} + 10^{-4}}.$$

Draw *K* vectors of counts for the three categories of events in the treatment arms:

$$\mathbf{Y}_{j,k} = (Y_{j,k1}, Y_{j,k2}, Y_{j,k3}) \sim \mathcal{M}\left(N_j^{(T)}, (\pi_{j,k1}^{(T)}, \pi_{j,k2}^{(T)}, \pi_{j,k3}^{(T)})\right)$$

Table 9. Pseudo-algorithm for generating simulation data for outcomes that have an is-subset-of relationship

Choose a combination of values for the parameters in Table 6.

Do 500 times:

Draw *K* vectors of counts for the three mutually exclusive categories implied by the two outcomes (comparator arms):

$$\mathbf{F}_{j,k} = (F_{j,k1}, F_{j,k2}, F_{j,k3}) \sim \mathcal{M}\left(N_j^{(C)}, (\gamma_{j,k1}^{(C)}, \gamma_{j,k2}^{(C)}, \gamma_{j,k3}^{(C)})\right).$$

If those experiencing outcome 1 are a subset of those experiencing outcome 2:

$A_{j,k1} = F_{j,k1}$, and $A_{j,k2} = F_{j,k1} + F_{j,k2}$.

Draw *K* vectors of study effects for outcomes 1 and 2

$(\theta_{j,k1}, \theta_{j,k2})' = \boldsymbol{\theta}_{j,k} \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{T}\right)$, where $\mathbf{T} = \mathrm{diag}(\boldsymbol{\tau})\,\mathbf{R}\,\mathrm{diag}(\boldsymbol{\tau})$

Calculate the probabilities for the three categories of events in the treatment arms:

$$\gamma_{j,k1}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k1} + \mathrm{logit}(\gamma_{j,k1}^{(C)})\right),$$

$$\gamma_{j,k1}^{(T)} + \gamma_{j,k2}^{(T)} = \mathrm{logit}^{-1}\left(\theta_{j,k2} + \mathrm{logit}(\gamma_{j,k1}^{(C)} + \gamma_{j,k2}^{(C)})\right),$$ with

$$\gamma_{j,k3}^{(T)} = 1 - \gamma_{j,k1}^{(T)} - \gamma_{j,k2}^{(T)}.$$

If $\gamma_{j,k3}^{(T)} \leq 0$, set

$$\gamma_{j,k3}^{(T)} := \frac{10^{-4}}{\gamma_{j,k1}^{(T)} + \gamma_{j,k2}^{(T)} + 10^{-4}},$$

$$\gamma_{j,k2}^{(T)} := \frac{\gamma_{j,k2}^{(T)}}{\gamma_{j,k1}^{(T)} + \gamma_{j,k2}^{(T)} + 10^{-4}},$$ and

$$\gamma_{j,k1}^{(T)} := \frac{\gamma_{j,k1}^{(T)}}{\gamma_{j,k1}^{(T)} + \gamma_{j,k2}^{(T)} + 10^{-4}}.$$

Draw *K* vectors of counts for the three mutually exclusive categories implied by the two outcomes (treatment arms):

$$\mathbf{G}_{j,k} = (G_{j,k1}, G_{j,k2}, G_{j,k3}) \sim \mathcal{M}\left(N_j^{(T)}, (\gamma_{j,k1}^{(T)}, \gamma_{j,k2}^{(T)}, \gamma_{j,k3}^{(T)})\right).$$

Then $B_{j,k1} = G_{j,k1}$ and $B_{j,k2} = G_{j,k1} + G_{j,k2}$.

Results

The results are organized by aim.

Description of the database (aim 1)

Out of the 4848 reviews in the Cochrane Library, 1919 had at least one collection of studies where the primary Cochrane reviewers performed a meta-analysis of binary data. 1381 candidate pairs of binary meta-analyses with at least seven RCTs reporting on both outcomes and having at least two events and 10 patients per arm were identified in 98 reviews. The vast majority did not fulfill eligibility criteria. The most common reason was that the counts for combinations of outcomes were not exactly recoverable. In the end 29 reviews contributed 45 eligible independent pairs or triplets of binary meta-analyses: 38 instances pertained to pairs of outcomes where the first outcome was a subset of the second outcome (Table 10, topic numbers 1 through 38); in five instances the pair of outcomes were mutually exclusive (Table 11, topic numbers 39 through 43); and for two topics we identified two triplets of outcomes with an is-subset-of relationship (Table 12, topic numbers 44 and 45). A single Cochrane review on medications for preventing post-operative nausea and vomiting contributed 12 independent pairs of outcomes that have an is-subset-of relationship.

The 45 topics pertain to a variety of clinical questions in cardiology (n=3), general medicine (n=2), pediatrics (n=6), obstetrics and gynecology (n=10), infectious diseases (n=5), nephrology (n=4), mental health (n=2) and surgery (n=13). In total, 1473 RCTs (258,675 randomized patients) were included in the database. The median number of RCTs per topic was 18, ranging from 7 to 203. The median total number of participants per topic was 3733 (range from 304 to 38,923). 21 topics had at least 20 RCTs, and 7 topics had at least 50 RCTs. Across the 45 topics, between 62% and 100% of RCTs reported on all outcomes. In 9 topics all RCTs reported data for all outcomes.

Table 10. Thirty eight pairs of outcomes having an is-subset-of relationship

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1 ID#*	O1 Description	O2 ID#*	O2 Description	O1 & O2 *K* (*N*)	Only O1 *K* (*N*)	Only O2 *K* (*N*)
1	Card	CD004587 (1)[11]	Drug-eluting stents versus bare metal stents for angina or acute coronary syndromes	Paclitaxel-eluting stents versus bare metal stents	1, 2	Major adverse cardiac events by 6 months	2, 2	Major adverse cardiac events by 12 months	15 (6325)	6 (1978)	1 (605)
2	Card	CD004587 (1)[11]	Drug-eluting stents versus bare metal stents for angina or acute coronary syndromes	Sirolimus-eluting stents versus bare metal stents	1, 1	Major adverse cardiac events by 6 months	2, 1	Major adverse cardiac events by 12 months	16 (4521)	5 (1141)	4 (1084)
3	Gen	CD002296 (5)[12]	Prevention of NSAID-induced gastroduodenal ulcers	Misoprostol (800 µg/day) versus placebo	8, 3	Dropout due to side events	9, 3	Dropout for any reason	10 (11798)	2 (1095)	1 (277)
4	Ped	CD001145 (6)[13]	Late postnatal corticosteroids for chronic lung disease in preterm infants	Late postnatal corticosteroids versus control	9, 1	Cerebral palsy	11, 1	Death or cerebral palsy	11 (777)	0	0
5	Ped	CD000509 (1)[14]	Inhaled nitric oxide for respiratory failure in preterm infants	Nitric oxide versus control	3, 1	Bronchopulmonary dysplasia among survivors at 36 weeks	4, 1	Death or bronchopulmonary dysplasia at 36 weeks	8 (958)	0	0
6	Ped	CD000551 (3)[15]	Ursodeoxycholic acid for primary biliary cirrhosis	Ursodeoxycholic acid versus nothing	1, 1	Mortality	2, 1	Mortality or liver transplantation	15 (1275)	0	1 (28)
7	Ped	CD001533 (1)[16]	Corticosteroid therapy for nephrotic syndrome in children	Increased versus standard dose of prednisone	4, 3	Number with frequent relapses	2, 3	Number with any relapse	7 (512)	0	0
8	Ped	CD003665 (1)[17]	Vitamin E for prevention of morbidity and mortality in preterm infants	Vitamin E versus placebo/no vitamin E	37, 1	Severe retrolental fibroplasia/retinopathy of prematurity	33, 1	Any retrolental fibroplasia/retinopathy of prematurity	8 (1666)	0	1 (51)
9	Ob Gyn	CD001396 (1)[18]	SSRIs for premenstrual syndrome	SSRIs versus placebo	11, 1	Withdrawal due to adverse events	11, 2	Withdrawal for any reason	32 (3486)	0	5 (247)
10	Ob Gyn	CD001750 (1)[19]	GnRH antagonists for assisted reproductive technology	GnRH antagonist versus long course GnRH agonist	2, 1	Ongoing pregnancy	3, 1	Clinical pregnancy	45 (7209)	4 (638)	17 (2195)
11	Ob Gyn	CD005214 (1)[20]	Depot medroxyprogesterone versus norethisterone oenanthate for long-acting progestogenic contraception.	Depot medroxyprogesterone versus norethisterone oenanthate	1, 1	Discontinuation at 12 months	1, 2	Discontinuation at 24 months	14 (2776)	1 (400)	0
12	Ob Gyn	CD004454 (1)[21]	Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth	Corticosteroids versus placebo/no treatment	6, 1	Neonatal deaths	4, 1	Fetal or neonatal deaths	18 (4140)	5 (513)	0
13	Ob Gyn	CD001141 (3)[22]	Support for breastfeeding mothers	All forms of breastfeeding support versus usual care	1, 3	Stopping any breastfeeding by 3 months	1, 5	Stopping any breastfeeding by 6 months	16 (5304)	4 (1247)	2 (664)
14	Ob Gyn	CD001141 (6)[22]	Support for breastfeeding mothers	Professional breastfeeding support versus usual care	1, 3	Stopping any breastfeeding by 3 months	1, 6	Stopping any breastfeeding by 6 months	9 (3539)	1 (507)	1 (849)

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1 ID#*	O1 Description	O2 ID#*	O2 Description	O1 & O2 *K* (*N*)	Only O1 *K* (*N*)	Only O2 *K* (*N*)
15	Ob Gyn	CD000014 (1)[23]	Amnioinfusion for meconium-stained liquor in labour	Amnioinfusion versus no amnioinfusion	6, 1	Caesarean for fetal distress	7, 1	Caesarean overall	11 (3380)	0	3 (615)
16	InfD	CD002848 (2)[24]	Rotavirus vaccine for preventing diarrhoea	Rhesus rotavirus vaccines versus control	2, 1	Severe episodes of rotavirus diarrhoea	1, 1	All episodes of rotavirus diarrhoea	20 (13305)	0	9 (2699)
17	InfD	CD002848 (2)[24]	Rotavirus vaccine for preventing diarrhoea	Bovine rotavirus vaccines versus control	2, 2	Severe episodes of rotavirus diarrhoea	1, 2	All episodes of rotavirus diarrhoea	17 (5283)	0	7 (1640)
18	InfD	CD003774 (1)[25]	Antivirals for preventing CMV disease in solid organ transplant recipients	Antiviral prophylaxis versus placebo/no treatment	1, 2	CMV syndrome	1, 1	All symptomatic CMV disease	19 (1981)	0	8 (411)
19	InfD	CD002898 (1)[26]	Antivirals and other therapies for HSV epithelial keratitis	Acyclovir versus idoxuridine	10, 1	Healing by 7 days	10, 2	Healing by 14 days	9 (476)	0	1 (75)
20	InfD	CD002898 (3)[26]	Antivirals and other therapies for HSV epithelial keratitis	Interferon/nucleoside antiviral versus control	7, 1	Healing by 7 days	7, 2	Healing by 14 days	11 (606)	0	2 (138)
21	Neph	CD003897 (1)[27]	IL2Ra for kidney transplant recipients	IL2Ra versus placebo/no treatment	3, 6	Graft loss censored for death with functioning graft	4, 4	Acute rejection: clinically suspected or biopsy proven	30 (5582)	0	1 (44)
22	Neph	CD003897 (2)[27]	IL2Ra for kidney transplant recipients	IL2Ra versus antithymocyte globulin	3, 3	Graft loss censored for death with functioning graft	2, 3	Graft loss or death with a functioning graft	12 (1394)	0	0
23	Neph	CD003961 (1)[28]	Tacrolimus versus cyclosporin for kidney transplant recipients	Tacrolimus versus cyclosporin	6, 3	Mortality	5, 3	Total graft loss (with death)	14 (2604)	0	0
24	Neph	CD004293 (1)[29]	Immunosuppression for idiopathic membranous nephropathy in adults with nephrotic syndrome	Immunosuppressive therapy versus placebo/no treatment	2, 1	Dialysis or transplantation	3, 1	Dialysis, transplantation or death	10 (620)	0	0
25	MH	CD003197 (3)[30]	Low dosage TCA for depression	Low dosage TCA versus placebo	3, 2	Depression improved by 2 weeks	3, 3	Depression improved by 4 weeks	11 (735)	1 (25)	2 (235)
26	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Ondansetron versus placebo	1, 33	Nausea	3, 28	Nausea or vomiting	132 (16967)	53 (10238)	18 (990)
27	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Droperidol versus placebo	1, 17	Nausea	3, 14	Nausea or vomiting	102 (8305)	33 (2199)	22 (1086)
28	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Dexamethasone versus placebo	2, 14	Vomiting	3, 8	Nausea or vomiting	70 (5807)	24 (2276)	4 (173)
29	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Metoclopramide versus placebo	1, 29	Nausea	3, 25	Nausea or vomiting	72 (3685)	18 (978)	14 (684)
30	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Granisetron versus placebo	2, 24	Vomiting	3, 18	Nausea or vomiting	54 (4206)	20 (1554)	2 (93)
31	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Tropisetron versus placebo	4, 39	Rescue antiemetic	3, 40	Nausea or vomiting	28 (2484)	12 (1296)	4 (480)

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1 ID#*	O1 Description	O2 ID#*	O2 Description	O1 & O2 *K* (*N*)	Only O1 *K* (*N*)	Only O2 *K* (*N*)
32	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Dolasetron versus placebo	1, 15	Nausea	3, 12	Nausea or vomiting	14 (2864)	2 (1070)	1 (52)
33	Surg	CD004125 (1)[31]	Prevention of postoperative nausea and vomiting	Hyoscine versus placebo	1, 22	Nausea	3, 20	Nausea or vomiting	15 (1040)	7 (411)	1 (32)
34	Surg	CD004125 (3)[31]	Prevention of postoperative nausea and vomiting	Droperidol versus granisetron	2, 53	Vomiting	3, 42	Nausea or vomiting	24 (1008)	6 (264)	0
35	Surg	CD004125 (3)[31]	Prevention of postoperative nausea and vomiting	Droperidol versus metoclopramide	1, 47	Nausea	3, 46	Nausea or vomiting	32 (964)	9 (209)	5 (238)
36	Surg	CD004125 (3)[31]	Prevention of postoperative nausea and vomiting	Granisetron versus metoclopramide	2, 71	Vomiting	3, 57	Nausea or vomiting	14 (388)	3 (70)	1 (12)
37	Surg	CD004125 (3)[31]	Prevention of postoperative nausea and vomiting	Dexamethasone versus ondansetron	4, 15	Rescue antiemetic	3, 22	Nausea or vomiting	7 (304)	0	0
38	Surg	CD004603 (2)[32]	Perioperative ketamine for acute postoperative pain	Adverse effects	1, 1	Nausea	1, 3	Nausea or vomiting	26 (1283)	0	14 (702)

The number of people experiencing the first outcome is a subset of the number of people experiencing the second outcome.
Card: cardiology; CMV: cytomegalovirus; Gen: general medicine; GnRH: gonadotrophin-releasing hormone; HSV: herpes simplex virus; InfD: infectious disease; MH: mental health; Neph: nephrology; NSAID: nonsteroidal anti-inflammatory drugs; Ob Gyn: obstetrics/gynecology; Ped: pediatrics; SSRI: Selective serotinine reuptake inhibitors; Surg: surgery; TCA: Tricyclic antidepressants.
The following are column keys: Comp #: comparison number in the Cochrane review; *K*: number of studies; *N*: number of participants; O1|O2: outcome 1 | 2.
* ID#: The first number is the number of the outcome in the respective Cochrane review; the second number is the number of the subgroup.

Table 11. Five pairs of mutually exclusive outcomes

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1 ID#*	O1 Description	O2 ID#*	O2 Description	O1 & O2 *K* (*N*)	Only O1 *K* (*N*)	Only O2 *K* (*N*)
39	Gen	CD001431 (1)[33]	Decision aids for people facing health treatment or screening decisions	Decision aids versus usual care	4, 1	Patient controlled decision making	4, 3	Practitioner controlled decision making	11 (1928)	0	1 (171)
40	Ped	CD004000 (1)[34]	Intravenous immunoglobulin for Kawasaki disease in children	Intravenous immunoglobulin versus control	1, 1	Development of chest aortic aneurysms (0 to 30 days)	1, 2	Development of chest aortic aneurysms (31 to 60 days)	10 (1024)	0	0
41	Ob Gyn	CD003766 (3)[35]	Continuous support for women during childbirth	Continuous support versus usual care	3, 1	Spontaneous vaginal birth	4, 1	Caesarean birth	14 (13093)	0	1 (420)
42	Ob Gyn	CD004659 (1)[36]	Antiplatelet agents for primary prevention of preeclampsia and its complications	Antiplatelet versus control	12, 2	Perinatal deaths	12, 1	Stillbirths or miscarriages	31 (24514)	3 (6445)	16 (7964)
43	Ob Gyn	CD004667 (2)[37]	Midwife-led versus other models of care for childbearing women	Midwife-led versus other models of care	9, 2	Spontaneous vaginal birth	7, 2	Caesarean birth	9 (9183)	0	2 (971)

Outcomes 1 and 2 are mutually exclusive. Gen: general medicine; Ob Gyn: obstetrics/gynecology; Ped: pediatrics.
The following are column keys: Comp #: comparison number in the Cochrane review; *K*: number of studies; *N*: number of participants; O1|O2: outcome 1 | 2.
* ID#: The first number is the number of the outcome in the respective Cochrane review; the second number is the number of the subgroup.

Table 12. Two triplets of outcomes having an is-subset-of relationship

#	Cate-gory	Review (Comp #)	Topic	Comparison	O1 ID#*	O1 Description	O2 ID#*	O2 Description	O3 ID#*	O3 Description
44	MH	CD001026 (1)[38]	Antidepressants plus benzodiazepines for major depression	Antidepressants plus benzodiazepines versus antidepressants	7, 1	Depression improved by 1 week	7, 2	Depression improved by 2 weeks	7, 3	Depression improved by 4 weeks
45	Card	CD002230 (5)[39]	Phosphodiesterase III inhibitors for heart failure	Phosphodiesterase III inhibitors versus control	3, 1	Sudden death	2, 1	Cardiac death	1, 1	Total mortality

The number of people experiencing the first outcome is a subset of the number of people experiencing the second outcome, which in turn is a subset of those experiencing the third outcome. MH: mental health; Card: cardiology.
The following are column keys: Comp #: comparison number in the Cochrane review; O1|O2|O3: outcome 1 | 2 | 3.
* ID#: The first number is the number of the outcome in the respective Cochrane review; the second number is the number of the subgroup.
For the first topic (number 44) a total of 10 RCTs with a total of 599 people provided data on all three outcomes. For the second topic (number 45), 13 RCTs (7337 people) provided data for all three outcomes, two RCTs (469 people) provided data only for the first and the third outcome, one RCT (230 people) provided data only for the second and the third outcome, and three RCTs (262 people) provided data only for the third outcome.

Empirical comparisons (aim 2)

We first describe results with meta-analysis models that use the normal approximation. Subsequently, we present the corresponding analyses with models that use the binomial or multinomial distribution to model the underlying data.

Normal approximation methods

Corrections for singularity in covariance matrices

Eighty studies in 21 topics, all in outcomes having an is-subset-of-relationship, had singular covariance matrices. Based on the formulas in the Appendix, it is obvious that singular covariance matrices can arise when the number of people experiencing the first outcome is exactly the same as the number experiencing the second outcome. For example, in topic 2, comparison between sirolimus-eluting and bare metal stents for acute cardiac disease, in the STRATEGY trial, the number of major cardiac events was the same at 6 months and 12 months, that is, no additional events were observed between 6 and 12 months.[40,41] The multivariate analyses used a ridge regression approach when at least one study had a singular covariance matrix. Table 13 shows the ridge regression factors (regularizing constants or regularizers) used. Using larger values or excluding the studies with singular covariance matrices from the analysis yields very similar results to those presented here (not shown). Figure 1 shows a plot of the meta-analysis means versus a range of values for the regularizing constant $\varepsilon_{\text{ridge}}$ for topic 2.

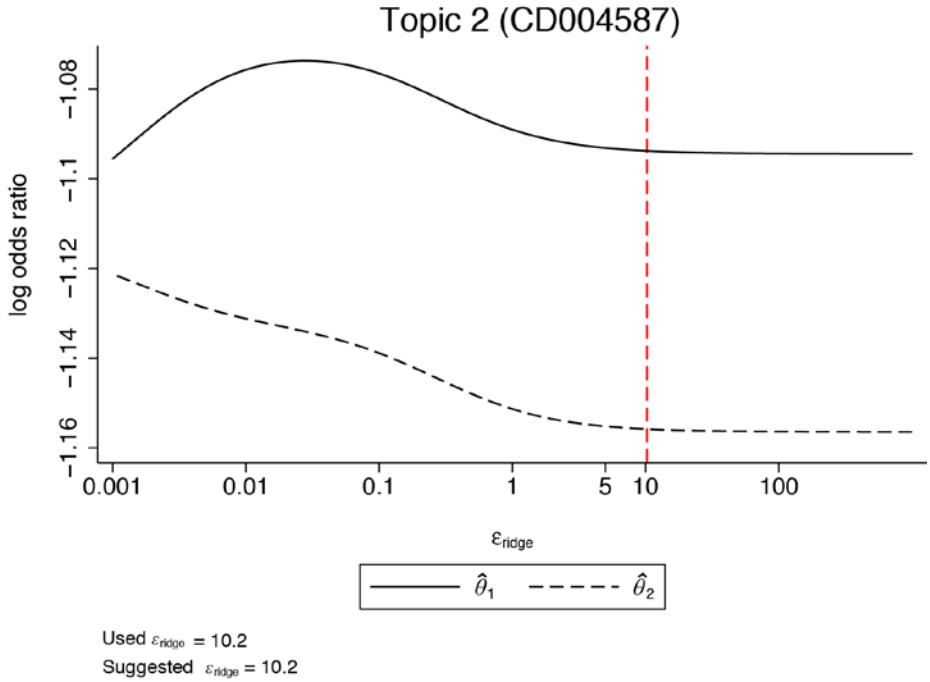
Table 13. Topics with RCTs with singular covariance matrices

Topic	*Number of RCTs with singular matrices*	*Number of RCTs reporting two or more outcomes*	*Total number of RCTs*	$\varepsilon_{\text{ridge}}$
2	1	16	25	10.2
3	1	10	13	15.8
4	1	11	11	2.5
6	5	15	16	4.8
9	2	32	37	2.5
10	1	45	66	23.4
12	1	18	23	6.2
22	2	12	12	4.3
23	1	14	14	2.7
24	4	10	10	4.4
26	18	132	203	70.8
27	14	102	157	77.6
28	1	70	98	7.6
29	10	72	104	32.4
31	1	28	44	4.1
32	3	14	17	10.0
33	1	15	23	3.2
34	1	24	30	3.4
35	4	32	46	6.5
38	7	26	40	14.8
45	1	16	17	575.4

The correction factors used in the ridge regression approach were selected using an objective heuristic (described in the Appendix). The remaining topics did not have any RCTs with singular covariance

matrices.

Figure 1. Plot for choosing the regularizing constant for the ridge regression (example, topic 2)

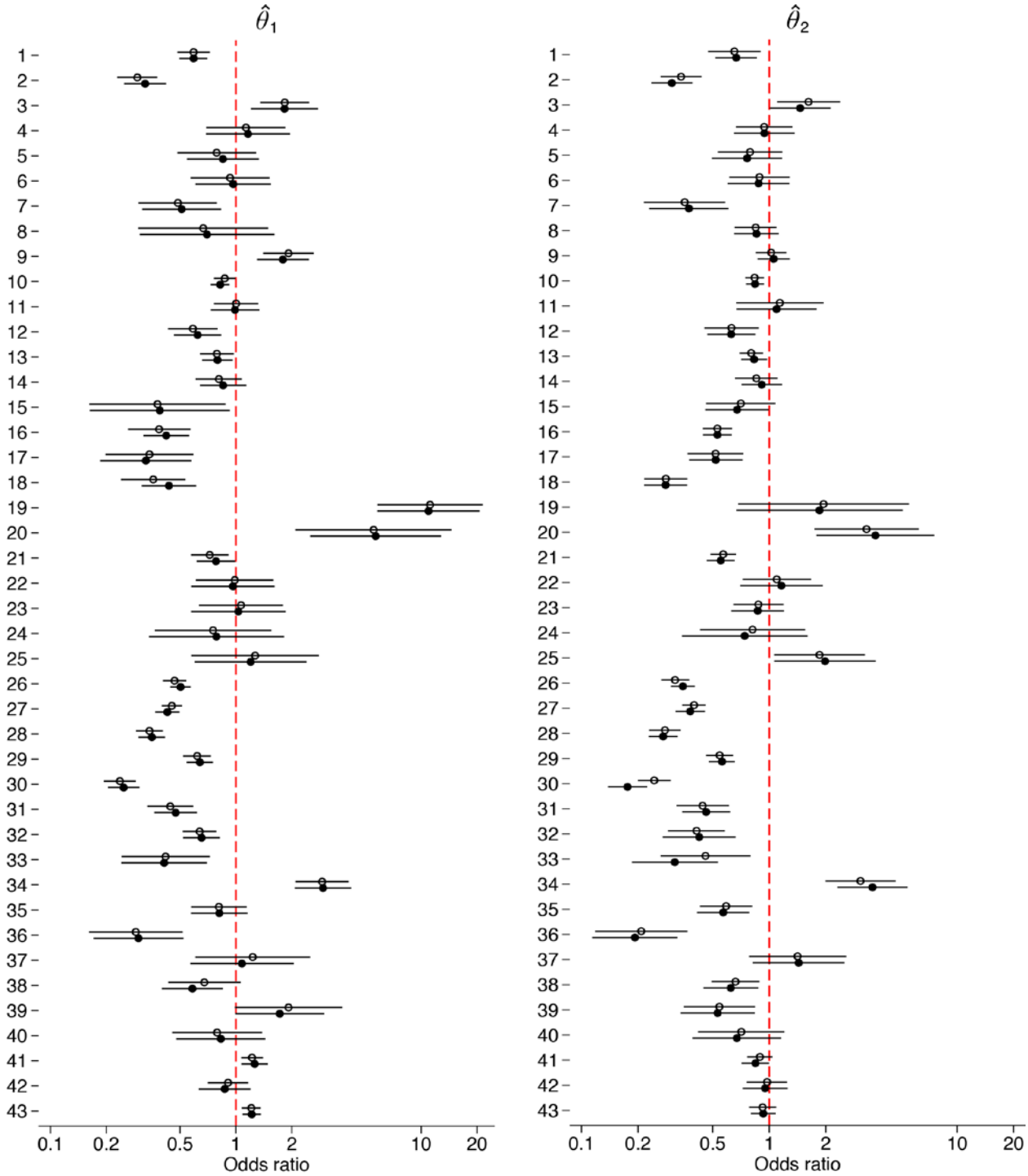


The suggested value of the regularizing constant uses a heuristic that examines the smoothed numerical derivatives with respect to $\varepsilon_{\text{ridge}}$ of the meta-analysis means and of the trace and determinant of the inverse covariance matrix for the means. An investigator [## redacted ##] examined such plots for all topics listed in Table 13; in all cases the $\varepsilon_{\text{ridge}}$ value suggested by the heuristic was deemed acceptable. An infinitely large value for $\varepsilon_{\text{ridge}}$ is equivalent to excluding studies with singular covariance matrices from the multivariate meta-analysis.

Figure 2 shows summary results with univariate and multivariate (bivariate) meta-analyses using the normal likelihood (in the Appendix we give the same results in tabular form). Figure 3 shows the respective results for topics 44 and 45, which have three outcomes with an is-subset-of relationship (the data in the figure are reported in tabular form in the Appendix). Overall, the point estimates and the lengths of the confidence intervals are comparable with the two methodologies.

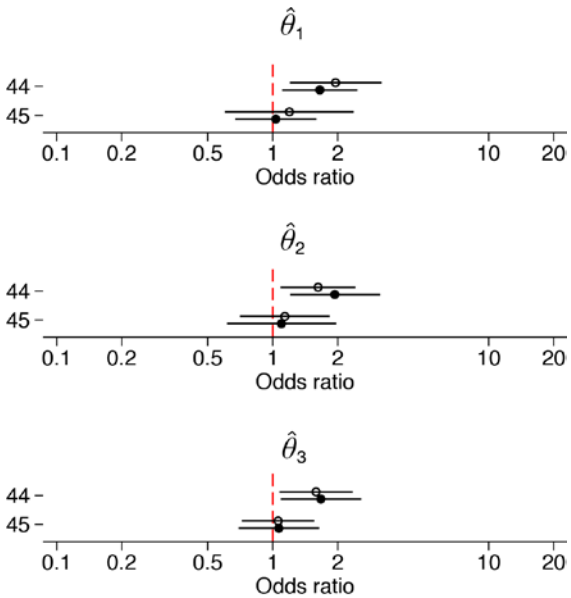
Some exceptions are notable. The most prominent is in topic 30, comparison of granisetron versus placebo for postoperative nausea and vomiting, where the summary odds ratio for “nausea or vomiting” (the second outcome) is more extreme in bivariate meta-analysis (0.18 95% CI 0.14 to 0.22) than in univariate meta-analysis (0.24, 95% CI 0.20 to 0.30). The summary odds ratios for the first outcome (“vomiting”) are 0.25 (0.20 to 0.30) and 0.24 (0.19 to 0.30), respectively. Nevertheless, even this numerical difference would not change conclusions on the effectiveness of granisetron. In all likelihood, systematic review conclusions that are based on the meta-analyses examined here would not change with the method of analysis.

Figure 2. Summary estimates from univariate and bivariate meta-analysis in topics 1 through 43 (REML random effects, normal approximation)



Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$: meta-analysis mean for the first outcome; $\hat{\theta}_2$: meta-analysis mean for the second outcome. For topics 1 through 38 those experiencing the first outcome are a subset of those experiencing the second outcome. For topics 39 through 43, the two outcomes are mutually exclusive.

Figure 3. Summary estimates from univariate and trivariate meta-analysis in topics 44 and 45 (REML random effects, normal approximation)



Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$ through $\hat{\theta}_3$: meta-analysis means for the first through third outcome.

As evident from Figure 2, Figure 3 and the Appendix tables, multivariate and univariate analyses yielded confidence intervals of different length. Across all 45 topics, the ratio of the standard errors with multivariate versus univariate analyses ranged between 0.6 and 1.4 for the first outcome and between 0.8 and 1.3 for the second outcome. The ratio was 1.1 for the third outcome in topics 44 and 45. Compared with univariate analyses, multivariate analyses yielded smaller standard errors (ratio of standard errors less than 1) in 25 and 22 topics for the first and second outcome, respectively.

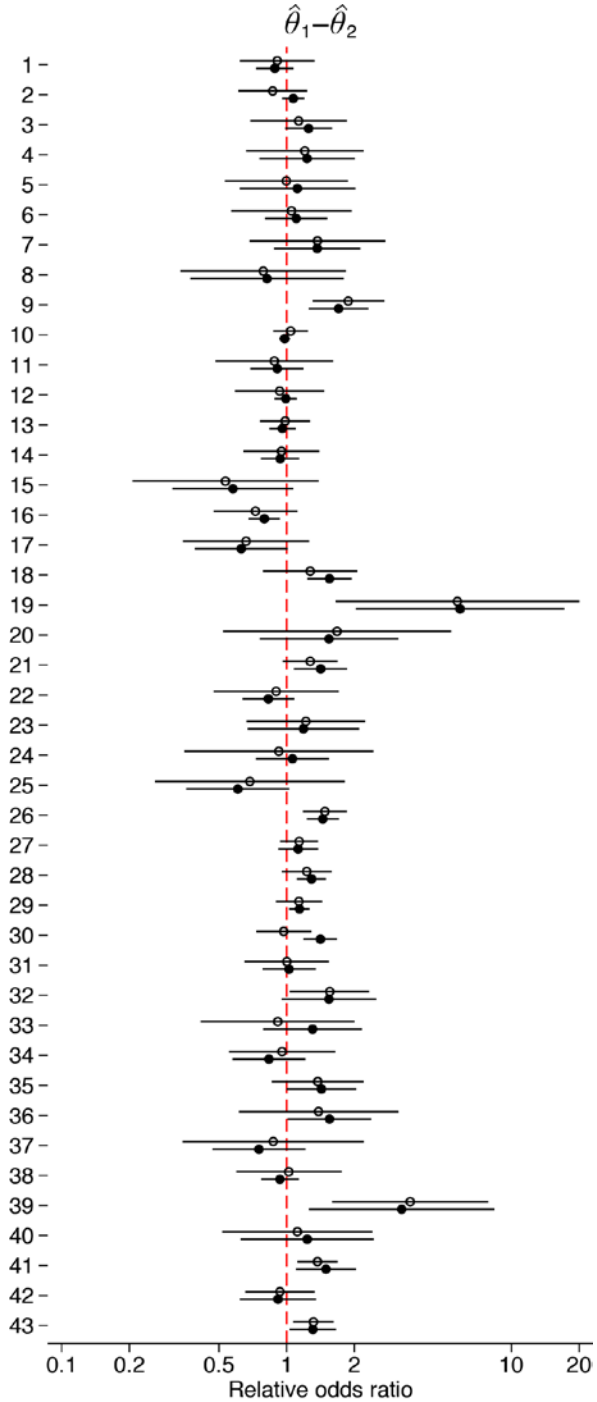
Between-outcome relative odds ratios (differences in treatment effects between outcomes)

It is instructive to calculate differences in the meta-analysis means. As discussed in the Methods section, such differences are log relative odds ratios for experiencing the first versus the second outcome in the treatment versus the comparator group. Depending on the context of the topic and the definitions of the outcomes, these differences may or may not be helpful. For example, if the first outcome is “alive at 12 months” and the second outcome is “alive at 6 months”, the difference in the outcomes can inform on whether the treatment log odds ratio differs between earlier and later time points. However, it is unclear whether the relative odds ratio is helpful or informative for pairs of outcomes such as “spontaneous vaginal birth” and “caesarean birth”. Regardless, forming differences between the outcomes presents an opportunity for technical observations and comments.

Figure 4 shows the relative treatment effects for univariate and bivariate meta-analyses in topics 1 through 43. Figure 5 shows the corresponding results for topics 44 and 45. Univariate and multivariate meta-analyses yield confidence intervals of different lengths. The reason is that univariate meta-analysis ignores the correlation between the meta-analysis effects, whereas multivariate meta-analysis does not. When this correlation is positive, the standard error of the

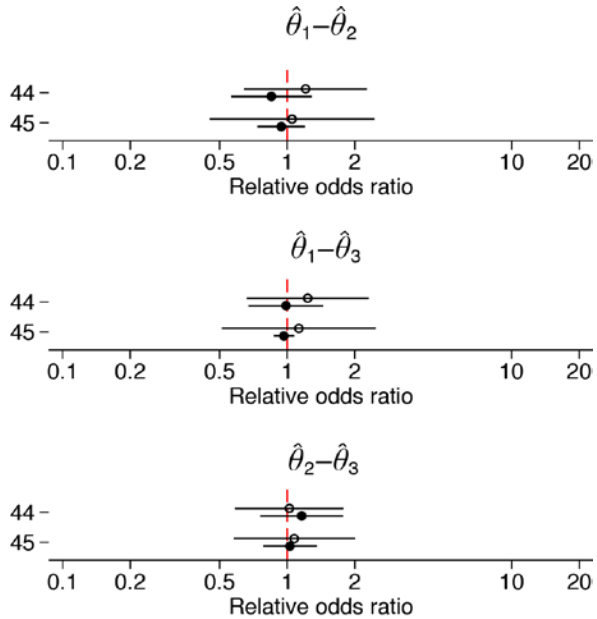
relative effect tends to be smaller in multivariate than in univariate meta-analysis; when the correlation is negative (as can happen for mutually exclusive outcomes), standard errors derived from multivariate analyses tend to be larger than those derived from univariate analyses (Figure 6).

Figure 4. Relative summary odds ratios from univariate and bivariate meta-analyses in topics 1 through 43



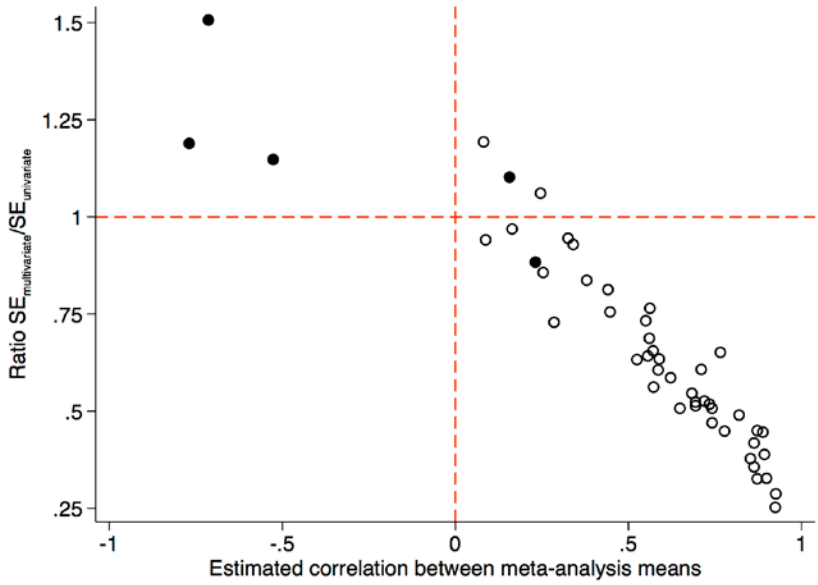
Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$ and $\hat{\theta}_2$: meta-analysis means for the first and second outcome, respectively.

Figure 5. Treatment effects with univariate and trivariate meta-analysis in topics 44 and 45.



Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. $\hat{\theta}_1$ through $\hat{\theta}_3$: meta-analysis means for the first through third outcome.

Figure 6. Ratio of standard errors for log relative odds ratios from multivariate and univariate meta-analysis and estimated correlation coefficient between means



Filled circles denote topics where the jointly meta-analyzed outcomes are mutually exclusive. Empty circles denote topics where outcomes have an is-subset-of relationship.

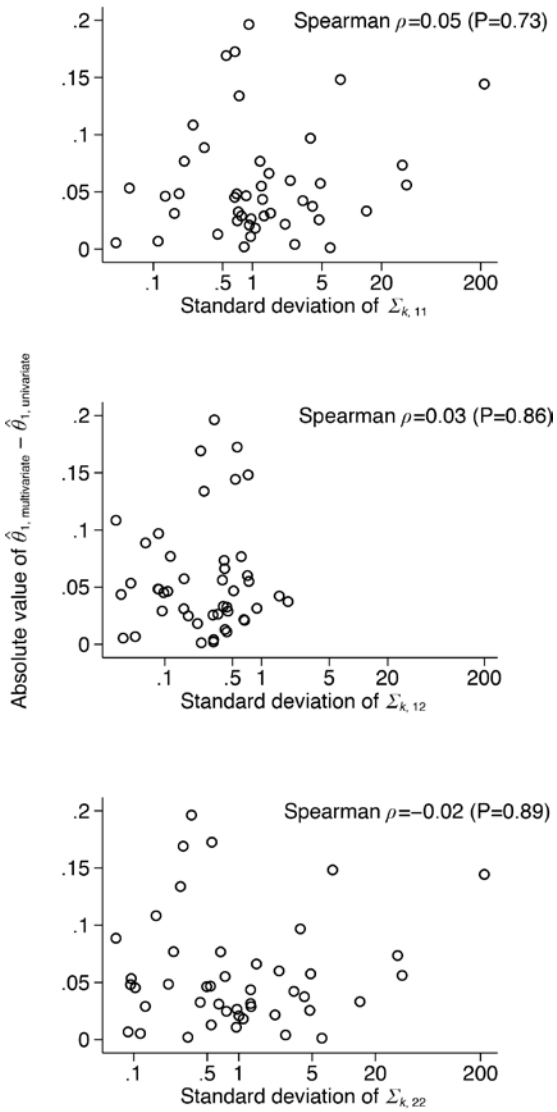
Information flow across outcomes in multivariate meta-analysis

Riley et al. provide formulas for the point estimates of the means and the standard errors for random effects bivariate meta-analysis in a restricted iterative generalized least squares (RIGLS) framework.[42] Upon inspection of the formulas it is evident that when the two outcomes are analyzed jointly, the point estimate of the first outcome depends on the study results for the second outcome if the following three conditions hold (a) the within-study covariances $\Sigma_{k,12}$ are nonzero; (b) the within-study variances $\Sigma_{k,11}$ are not all equal ($\Sigma_{k,11} \neq \Sigma_{11}$); and (c) the within-study covariances $\Sigma_{k,12}$ are not all equal ($\Sigma_{k,12} \neq \Sigma_{12}$). Similarly, the point estimate of the second outcome depends on the study results for the first outcome if a set of analogous conditions holds. (The only change is in condition (b), which becomes $\Sigma_{k,22} \neq \Sigma_{22}$). The covariance matrices for the summary effects always depend on both outcomes.

If information flows from the second outcome to the first outcome, the point estimates of the first outcome may be different under the univariate and multivariate meta-analysis. Reciprocally, if information flows from the first outcome to the second outcome, the point estimates of the second outcome may be different with univariate and multivariate meta-analysis. Figure 7 and Figure 8 examine the relationship between the absolute difference in the meta-analysis estimates with univariate and multivariate methods, and the standard deviation of the within-study variances and covariances across the 45 topics. No significant association is observed (P-values for Spearman's correlation are all 0.27 or larger).

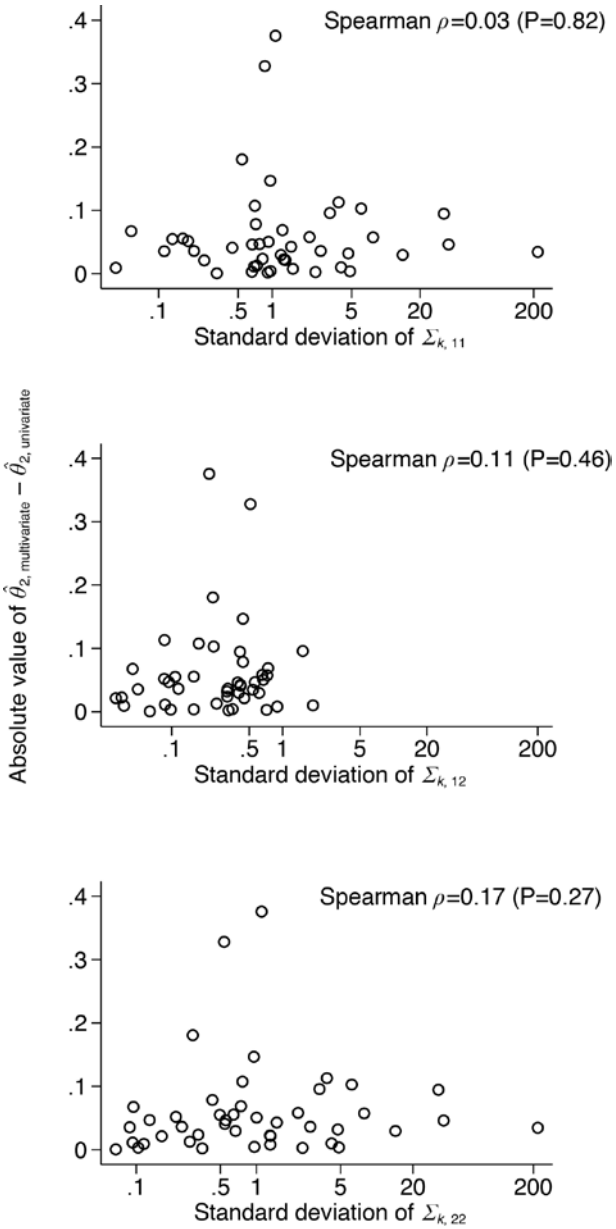
Finally, information can flow towards the first meta-analysis mean, when additional studies exist for the second outcome, and reciprocally. Figure 9 provides scatter plots of the absolute difference in the meta-analysis estimates with multivariate and univariate models and the number of studies reporting only the other outcome. Overall, the observed association is not beyond what is expected by chance.

Figure 7. Difference in the point estimate for the first outcome with univariate and multivariate meta-analysis versus the variability in the elements of the within-study covariance matrix (topics 1 through 45)



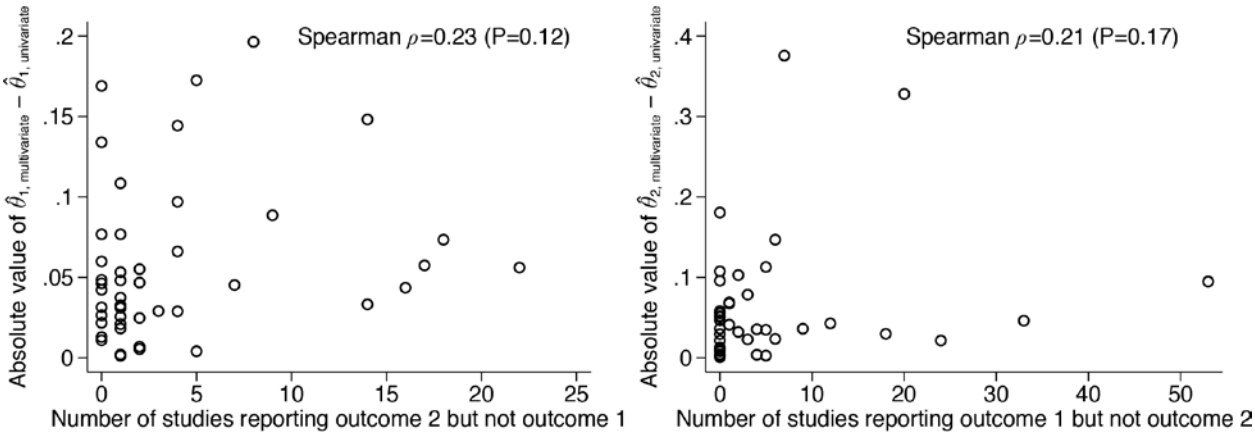
Differences between univariate and multivariate point estimates for the first outcome suggest flow of information from the second outcome to the first outcome.

Figure 8. Difference in the point estimate for the second outcome with univariate and multivariate meta-analysis versus the variability in the elements of the within-study covariance matrix (topics 1 through 45)



Differences between univariate and multivariate point estimates for the second outcome suggest flow of information from the first outcome to the second outcome.

Figure 9. Difference in the point estimates with univariate and multivariate meta-analysis versus the number of studies reporting the other outcome (topics 1 through 45)



Estimation of the between-study correlation parameters

Between-study correlations were poorly estimated in almost all topics (Table 14). The point estimates are approximately +1 or -1; the respective variances (obtained by applying the delta method to the fitted multivariate meta-analysis model results) were very large, indicating a relatively flat likelihood with respect to the correlation parameter. In some sense, the fitted models provide information only on the sign of the correlation. This observation is concordant with published analyses of single examples.[9]

Table 14. Estimated between-study correlation between the first and second outcome

Topic	$\hat{\rho}_{12}$ (var($\hat{\rho}_{12}$))	*Topic*	$\hat{\rho}_{12}$ (var($\hat{\rho}_{12}$))	*Topic*	$\hat{\rho}_{12}$ (var($\hat{\rho}_{12}$))
1	1.00 (-)	16	1.00 (-)	31	1.00 (-)
2	1.00 (-)	17	0.63 (0.10)	32	-1.00 (-)
3	1.00 (-)	18	-0.78 (-)	33	1.00 (-)
4	-1.00 (-)	19	1.00 (-)	34	1.00 (-)
5	-1.00 (-)	20	1.00 (-)	35	0.08 (-)
6	1.00 (-)	21	-1.00 (-)	36	0.05 (-)
7	1.00 (-)	22	1.00 (-)	37	0.99 (-)
8	1.00 (-)	23	-0.43 (2.59)	38	-1.00 (-)
9	-1.00 (-)	24	1.00 (-)	39	-1.00 (-)
10	-1.00 (0.04)	25	0.91 (0.13)	40	1.00 (-)
11	1.00 (-)	26	0.13 (0.95)	41	-1.00 (-)
12	1.00 (-)	27	1.00 (-)	42	1.00 (-)
13	1.00 (-)	28	1.00 (-)	43	-0.99 (-)
14	1.00 (-)	29	0.21 (-)	44	-1.00 (-)
15	1.00 (-)	30	1.00 (-)	45	1.00 (-)

The dash ("-") denotes an estimated variance larger than 1000. In the trivariate meta-analyses in topics 44 and 45, the estimated correlation coefficients (and variances) between the first and third outcome were -1 (42.2) and 0.98 (0.02), respectively; and between the second and third outcome they were 1 (>1000) and 0.97 (>1000).

Meta-analyses using the binomial or multinomial likelihood

Analyses with these models were performed in the Bayesian framework (see "Meta-analysis models using the discrete likelihood", page 18). Modeling within-study variation based on the binomial (for univariate meta-analyses) or the multinomial (for multivariate meta-analyses) distribution does not require continuity corrections or the use of a regularizer (a ridge regression correction). Overall, results were very similar to those obtained from models using the normal approximation. For example, Figure 10 and Figure 11 are the counterparts of Figure 2 and Figure 3, respectively. Similarly, Figure 12 and Figure 13 are the counterparts of Figure 4 and Figure 5. The numerical results are provided in the Appendix.

The posterior distribution of the between-study correlation parameter(s) was (were) very sensitive to the choice of the prior (Table 15). This is congruent with the findings in Table 14 for analyses using the normal approximation. There was agreement in the sign of the correlation coefficients between analyses using the normal approximation (frequentist analyses) and analyses using the discrete likelihood (Bayesian analyses, uninformative priors). The kappa for agreement was 0.42, P=0.001. However, the difference is that the estimates for the coefficients have moved to values other than -1 or +1.

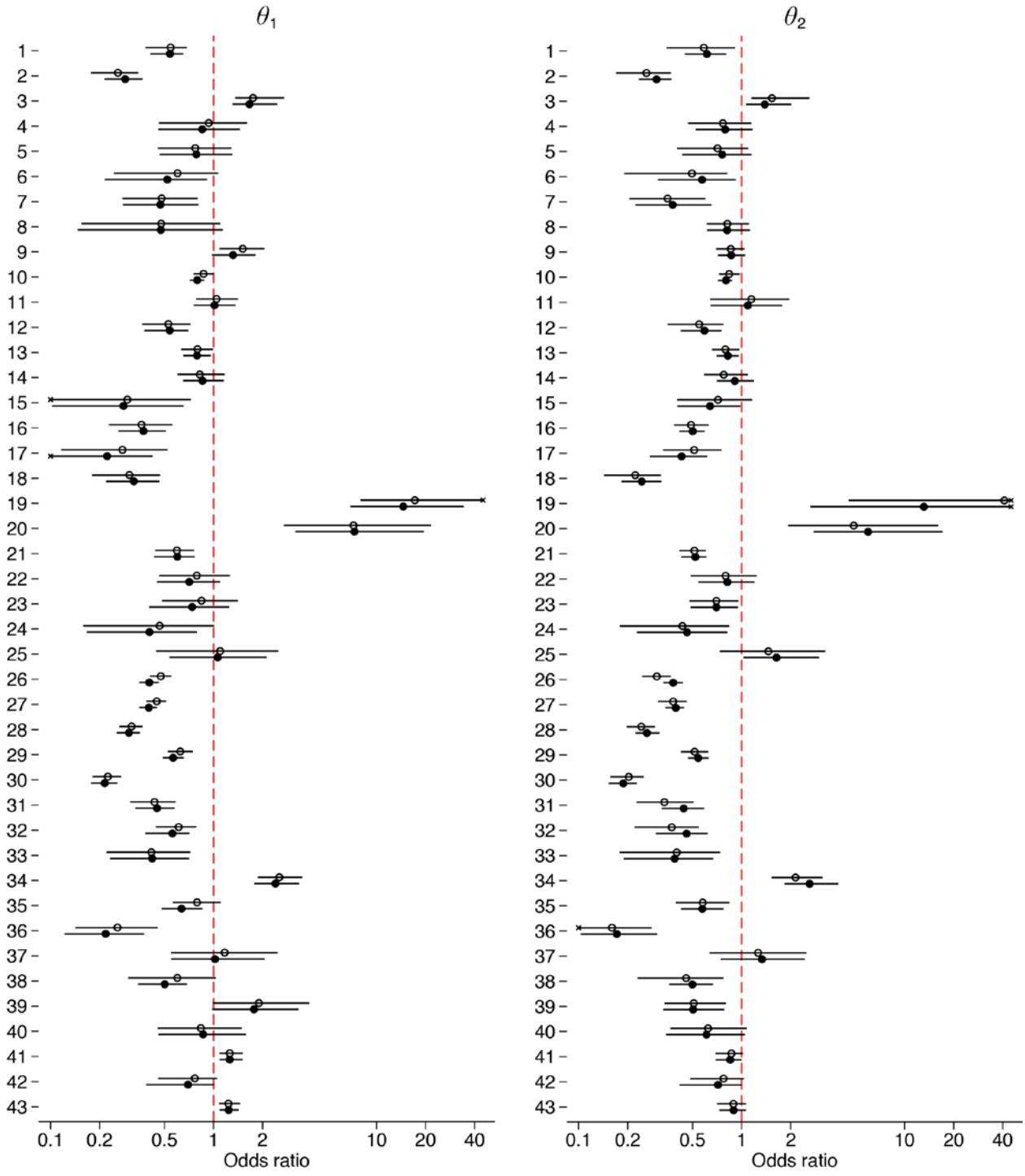
Table 15. Posterior median and 95% credible interval for between-study correlations with uninformative priors and priors informative on the correlation sign

Quantity or topic	Uninformative prior	Prior informative on the sign of the correlation
ρ_{12}		
1	0.42 (-0.82, 0.96)	0.69 (-0.66, 0.98)
2	0.45 (-0.88, 0.98)	0.82 (-0.58, 0.99)
3	0.29 (-0.91, 0.97)	0.68 (-0.73, 0.98)
4	0.28 (-0.86, 0.96)	0.64 (-0.75, 0.98)
5	0.08 (-0.93, 0.94)	0.48 (-0.84, 0.98)
6	0.76 (-0.68, 0.99)	0.90 (-0.26, 0.99)
7	0.19 (-0.92, 0.96)	0.66 (-0.78, 0.99)
8	0.04 (-0.94, 0.94)	0.57 (-0.80, 0.99)
9	-0.03 (-0.95, 0.92)	0.32 (-0.86, 0.96)
10	0.40 (-0.85, 0.97)	0.77 (-0.70, 0.99)
11	0.91 (0.27, 0.99)	0.94 (0.54, 0.99)
12	0.87 (-0.11, 0.99)	0.92 (0.09, 0.99)
13	0.52 (-0.81, 0.98)	0.80 (-0.57, 0.99)
14	0.66 (-0.80, 0.98)	0.87 (-0.58, 0.99)
15	0.68 (-0.32, 0.98)	0.84 (-0.02, 0.99)
16	0.90 (0.36, 0.99)	0.93 (0.54, 0.99)
17	0.63 (-0.27, 0.93)	0.74 (-0.03, 0.95)
18	0.24 (-0.91, 0.96)	0.62 (-0.84, 0.98)
19	0.15 (-0.92, 0.95)	0.61 (-0.79, 0.98)
20	0.77 (-0.29, 0.98)	0.88 (0.12, 0.99)
21	-0.20 (-0.95, 0.83)	0.24 (-0.84, 0.96)
22	0.40 (-0.90, 0.97)	0.79 (-0.67, 0.99)
23	0.15 (-0.91, 0.92)	0.49 (-0.78, 0.97)
24	0.48 (-0.90, 0.98)	0.84 (-0.53, 0.99)
25	0.52 (-0.87, 0.97)	0.78 (-0.56, 0.99)
26	0.98 (0.93, 0.99)	0.99 (0.95, 1.00)
27	0.91 (-0.47, 0.99)	0.91 (-0.47, 0.99)

Quantity or topic	Uninformative prior	Prior informative on the sign of the correlation
28	0.78 (-0.58, 0.98)	0.89 (-0.02, 0.99)
29	0.43 (-0.92, 0.98)	0.80 (-0.70, 0.99)
30	0.75 (-0.82, 0.98)	0.88 (-0.18, 0.99)
31	0.55 (-0.81, 0.98)	0.85 (-0.48, 0.99)
32	0.27 (-0.89, 0.97)	0.64 (-0.75, 0.98)
33	0.71 (-0.62, 0.98)	0.88 (-0.15, 0.99)
34	-0.01 (-0.94, 0.89)	0.43 (-0.79, 0.99)
35	0.19 (-0.96, 0.97)	0.70 (-0.71, 0.99)
36	0.26 (-0.90, 0.96)	0.68 (-0.72, 0.99)
37	0.29 (-0.90, 0.97)	0.69 (-0.78, 0.99)
38	0.59 (-0.87, 0.98)	0.89 (-0.35, 0.99)
39	-0.80 (-0.98, 0.11)	Not run*
40	0.53 (-0.61, 0.97)	Not run*
41	-0.60 (-0.98, 0.71)	Not run*
42	0.73 (-0.45, 0.98)	Not run*
43	-0.20 (-0.96, 0.92)	Not run*
44	0.44 (-1.00, 1.00)	0.81 (0.07, 1.00)
45	0.99 (0.42, 1.00)	0.99 (0.18, 1.00)
ρ_{13}		
44	0.32 (-1.00, 1.00)	0.80 (0.06, 1.00)
45	0.98 (0.17, 1.00)	0.97 (-0.08, 1.00)
ρ_{23}		
44	0.16 (-0.96, 0.99)	0.78 (0.16, 1.00)
45	0.99 (0.72, 1.00)	0.99 (0.68, 1.00)

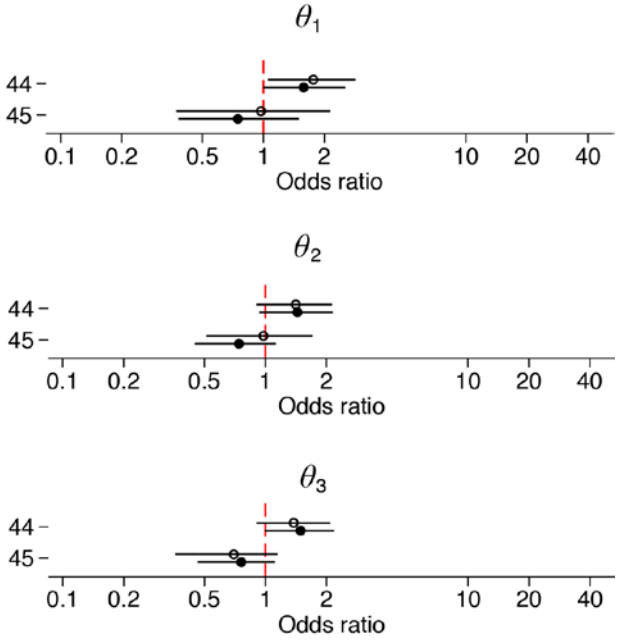
*Topics 39-43 correspond to mutually exclusive outcomes. We have no intuition on whether the between-study correlation among the two log odds ratios is expected to be positive or negative. Recall that although the proportions of people experiencing mutually exclusive outcomes are negatively correlated, the respective treatment effects can be positively or negatively correlated.

Figure 10. Comparison of univariate and multivariate meta-analysis using the binomial or the multinomial distribution to model within-study variance (topics 1 through 43)



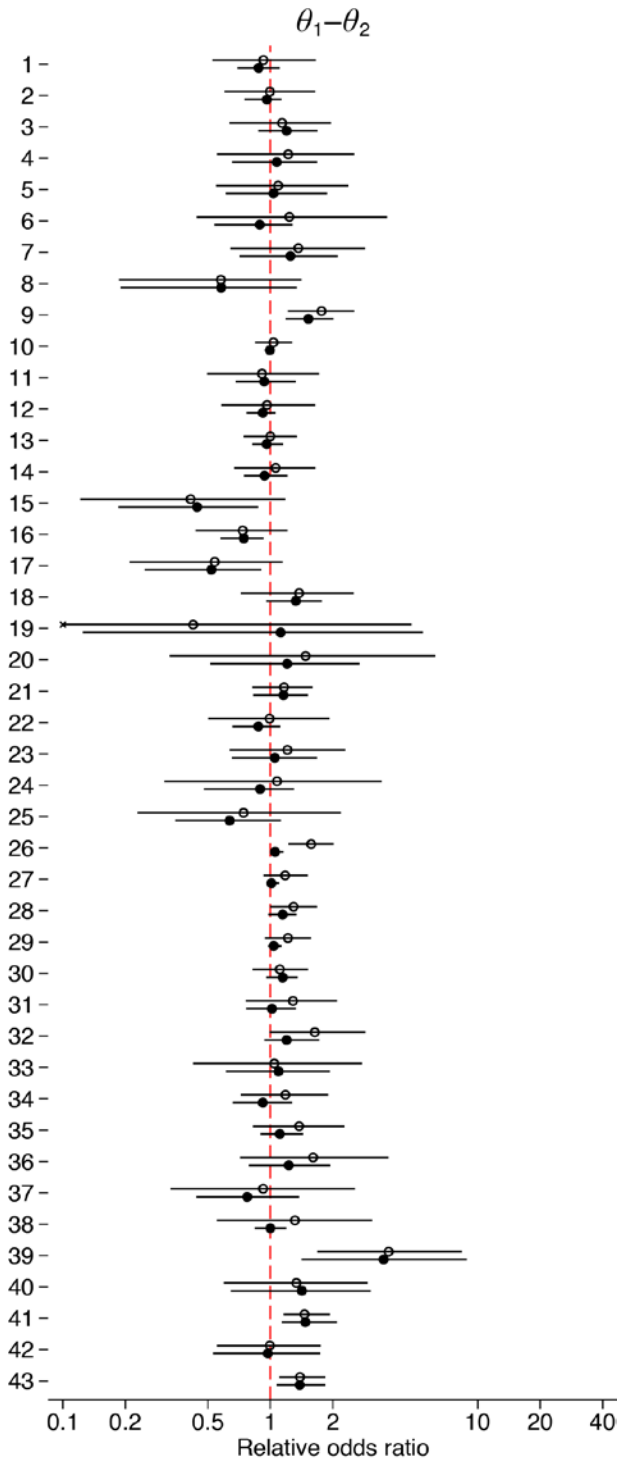
Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1: meta-analysis posterior median for the first outcome; θ_2: meta-analysis posterior median for the second outcome. For topics 1 through 38 those experiencing the first outcome are a subset of those experiencing the second outcome. For topics 39 through 43, the two outcomes are mutually exclusive. Small "x" markers denote truncated credible intervals.

Figure 11. Comparison of univariate and trivariate meta-analysis using the binomial or the multinomial distribution to model within-study variance (topics 44 and 45)



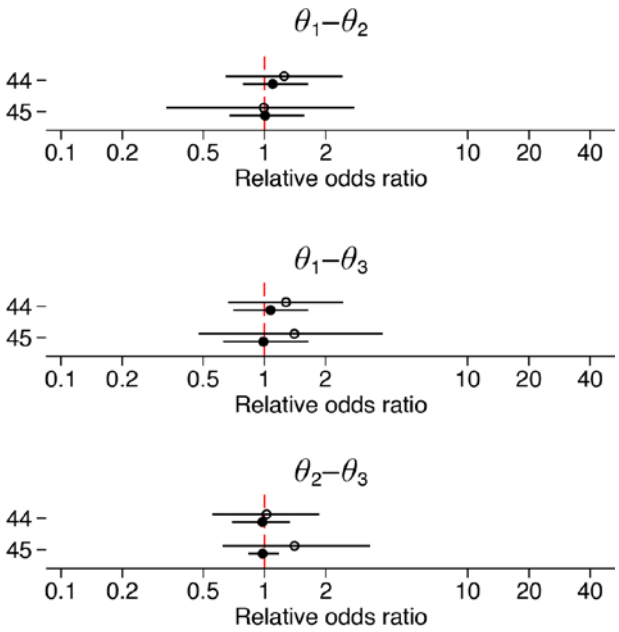
Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 through θ_3: meta-analysis posterior median for the three outcomes. The three outcomes have an is-subset-of relationship.

Figure 12. Relative summary odds ratios from univariate and bivariate meta-analyses using the binomial or the multinomial distribution to model within-study variance (topics 1 through 43)



Filled circles are results from bivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 and θ_2: meta-analysis posterior medians for the first and second outcome, respectively. For topics 1 through 38 those experiencing the first outcome are a subset of those experiencing the second outcome. For topics 39 through 43, the two outcomes are mutually exclusive.Small “x” markers denote truncated credible intervals.

Figure 13. Relative summary odds ratios from univariate and trivariate meta-analyses using the binomial or the multinomial distribution to model within-study variance (topics 44 and 45)



Filled circles are results from trivariate meta-analyses. Empty circles are results from univariate meta-analyses. θ_1 through θ_3: meta-analysis posterior median for the three outcomes. The three outcomes have an is-subset-of relationship.

Illustrative simulation analyses (aim 3)

Fidelity of simulations and code integrity

The means and covariances of the simulated counts in the comparator arm and the true effects $\theta_{k,1}$ and $\theta_{k,2}$ matched the simulation parameters closely for all 648 scenarios for mutually exclusive outcomes and all 648 scenarios for outcomes with an is-subset-of relationship. Detailed inspection of graphs of the distribution of the above parameters in scenarios corresponding to combinations of highest and lowest values for parameters, as well as 15 additional randomly chosen scenarios were not suggestive of any systematic errors in programing and analysis.

Exploration of influential simulation parameters with analysis of variance (ANOVA)

We used ANOVA to identify influential simulation parameters for the MSE, bias and coverage probability under univariate and multivariate meta-analysis. ANOVA was used as a screening tool. We examined up to two-way interactions between the most influential parameters. Table 16 is for simulations of mutually exclusive outcomes. Note that these simulations and their conclusions are symmetric with respect to the two outcomes, in that findings for the second outcome are analogous to findings for the first outcome. Table 17 outlines corresponding explorations for the case of outcomes that have an is-subset-of relationship. Note that the patterns of important factors differ between the first and the second outcome.

Why are the MSE, bias and coverage dependent on μ_1 or μ_2?

Interestingly, the MSE, bias and coverage depend on μ_1 or μ_2. The reason may not be immediately obvious; if anything, one would expect these quantities to be independent of the true means. This is because in a bivariate meta-analysis using the normal approximation, the partial derivative of the log likelihood with respect to the means is independent of the variances, and reciprocally, the partial derivative of the log likelihood with respect to the variances is independent of the means.

However, recall that in the normal approximation model, the conditional sample covariance matrices (the within-study covariance matrices) are calculated as functions of the proportions of events, and are thus dependent on the true means. Therefore, a correlation between the means and the calculated covariance matrices exists. This situation is analogous to univariate meta-analysis, where for non-variance-stabilizing link functions such as the log() or the logit() functions, the study effects (log risk ratio or log odds ratio, respectively) are correlated with their variances. The multivariate meta-analysis methods that use the multinomial to model within-study variance do not have this shortcoming.

Table 16. Influential simulation parameters for random effects bivariate meta-analysis of two mutually exclusive outcomes

ANOVA factors	$\hat{\theta}_1$			$\hat{\theta}_2$		
	MSE	*Coverage*	*Bias*	*MSE*	*Coverage*	*Bias*
K	2	1	–	2	1	–
N	1	2	2	1	2	2
μ_1	4	–	1	–	–	–
μ_2	–	–	–	4	–	1
$K \times N$	3	3	–	3	3	–
$K \times \mu_1$	6	–	–	–	–	–
$K \times \mu_2$	–	–	–	6	–	–
$N \times \mu_1$	5	–	3	–	–	–
$N \times \mu_2$	–	–	–	5	–	3
τ_1^2	–	–	–	–	–	–
τ_2^2	–	–	–	–	–	–
ρ	–	–	–	–	–	–

The numbers stand for the ranking of factors in an ANOVA model in order of decreasing statistical significance. The dash stands for nonsignificant factors or interactions. Note that the ranks do not reflect the relative sizes of the mean squares in the ANOVA.

Table 17. Influential simulation parameters for random effects bivariate meta-analysis of two outcomes with an is-subset-of relationship

ANOVA factors	$\hat{\theta}_1$			$\hat{\theta}_2$		
	MSE	*Coverage*	*Bias*	*MSE*	*Coverage*	*Bias*
K	2	1	–	2	1	–
N	1	2	2	1	–	–
μ_1	4	4	1	–	–	–
μ_2	–	–	4	4	–	1
$K \times N$	3	–	6	3	–	–
$K \times \mu_1$	5	–	5	–	–	–
$K \times \mu_2$	–	–	–	6	–	–
$N \times \mu_1$	6	3	3	7	–	–
$N \times \mu_2$	–	–	7	5	–	–
τ_1^2	–	–	–	–	–	–
τ_2^2	–	–	–	–	–	–
ρ	–	–	–	–	–	–

Note that the ranks do not reflect the relative sizes of the mean squares in the ANOVA.

MSE, bias and coverage for bivariate random effects meta-analysis

Based on the explorative analyses in Table 16 and Table 17 we report the MSE, bias and coverage averaging over choices for between-study heterogeneity and between-study

correlations in the true means. Table 18 shows these results for two mutually exclusive outcomes. Table 19 shows the results for two outcomes having an is-subset-of relationship. Overall, the pattern of results is similar in the two tables. Note that the coverage is consistently above the desired 95%, which suggests that the t distribution with one degree of freedom is somewhat conservative for the scenarios examined here. Results for scenarios where half the studies reported both outcomes and half reported only the first outcomes were similar (not shown).

Table 18. MSE, coverage and bias for random effects bivariate meta-analysis (two mutually exclusive outcomes) averaging over τ_1^2, τ_2^2 and ρ

K	N	$\exp(\mu_1)$	$\exp(\mu_2)$	MSE for $\hat{\theta}_1$ $(\times 10^{-3})$	MSE for $\hat{\theta}_2$ $(\times 10^{-3})$	Bias for $\hat{\theta}_1$ $(\times 10^{-3})$	Bias for $\hat{\theta}_2$ $(\times 10^{-3})$	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
10	50	1	1	28	28	-2	0	98.8	98.8
10	50	1	1.5	28	25	0	-6	98.7	98.8
10	50	1.5	1	25	28	-5	-1	98.7	98.9
10	50	1.5	1.5	25	25	-4	-5	98.7	98.8
10	100	1	1	14	14	0	-1	98.6	98.9
10	100	1	1.5	14	13	0	-1	98.8	98.6
10	100	1.5	1	13	14	-2	-1	98.6	98.5
10	100	1.5	1.5	13	13	-1	-2	98.5	98.4
10	500	1	1	3	3	0	0	98.4	98.3
10	500	1	1.5	3	3	0	-1	98.6	98.5
10	500	1.5	1	3	3	0	1	98.4	98.5
10	500	1.5	1.5	3	3	0	0	98.5	98.3
20	50	1	1	14	14	0	1	98.1	97.9
20	50	1	1.5	14	13	-1	-9	98.0	97.7
20	50	1.5	1	13	13	-11	1	97.5	98.0
20	50	1.5	1.5	12	12	-7	-6	97.8	97.6
20	100	1	1	7	7	0	-1	97.5	97.6
20	100	1	1.5	7	6	1	-3	97.4	97.5
20	100	1.5	1	6	7	-3	2	97.4	97.4
20	100	1.5	1.5	6	6	-2	-2	97.5	97.4
20	500	1	1	1	1	0	0	97.4	97.4
20	500	1	1.5	1	1	1	-1	97.2	97.4
20	500	1.5	1	1	1	0	0	97.5	97.4
20	500	1.5	1.5	1	1	-1	0	97.3	97.1

Table 19. MSE, coverage and bias for random effects bivariate meta-analysis (two outcomes that have an is-subset-of relationship) averaging over τ_1^2, τ_2^2 and ρ

K	N	$\exp(\mu_1)$	$\exp(\mu_2)$	MSE for $\hat{\theta}_1$ $(\times 10^{-3})$	MSE for $\hat{\theta}_2$ $(\times 10^{-3})$	Bias for $\hat{\theta}_1$ $(\times 10^{-3})$	Bias for $\hat{\theta}_2$ $(\times 10^{-3})$	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
10	50	1	1	27	18	-2	-1	98.9	98.7
10	50	1	1.5	28	17	-1	1	98.7	98.6
10	50	1.5	1	25	18	-15	2	98.7	98.5
10	50	1.5	1.5	25	17	-11	2	98.7	98.6
10	100	1	1	14	9	0	-1	98.7	98.5
10	100	1	1.5	14	9	0	2	98.8	98.5
10	100	1.5	1	13	9	-7	-1	98.5	98.6
10	100	1.5	1.5	13	9	-5	1	98.6	98.5

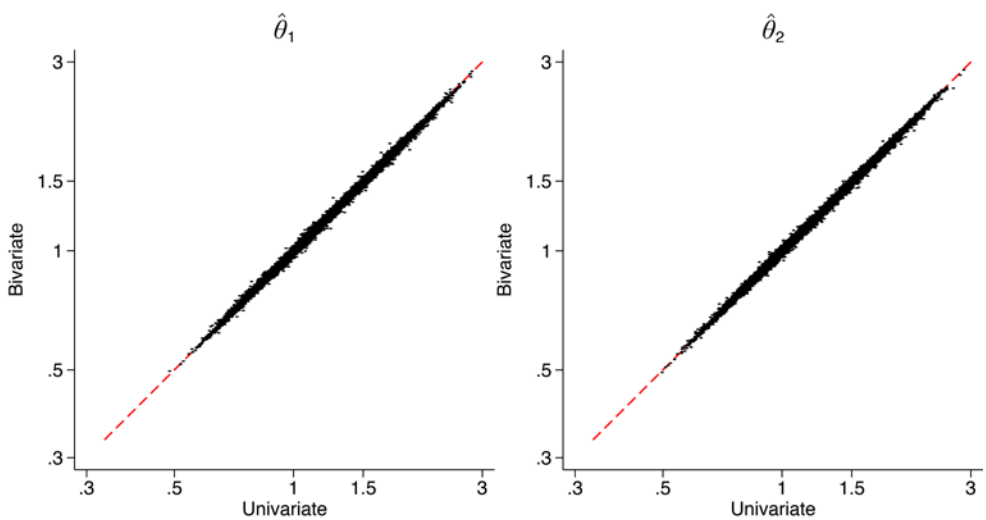
K	N	$\exp(\mu_1)$	$\exp(\mu_2)$	MSE for $\hat{\theta}_1$ $(\times 10^{-3})$	MSE for $\hat{\theta}_2$ $(\times 10^{-3})$	Bias for $\hat{\theta}_1$ $(\times 10^{-3})$	Bias for $\hat{\theta}_2$ $(\times 10^{-3})$	Coverage (%) for $\hat{\theta}_1$	Coverage (%) for $\hat{\theta}_2$
10	500	1	1	3	2	0	0	98.4	98.5
10	500	1	1.5	3	2	0	0	98.5	98.5
10	500	1.5	1	3	2	-2	0	98.5	98.5
10	500	1.5	1.5	3	2	-1	0	98.4	98.5
20	50	1	1	13	9	0	1	98.0	97.5
20	50	1	1.5	13	9	-2	-1	98.1	97.3
20	50	1.5	1	12	9	-23	0	97.3	97.4
20	50	1.5	1.5	12	9	-15	2	97.8	97.3
20	100	1	1	7	4	-1	-1	97.9	97.6
20	100	1	1.5	7	4	1	1	97.7	97.4
20	100	1.5	1	6	4	-10	1	97.4	97.4
20	100	1.5	1.5	6	4	-6	0	97.6	97.4
20	500	1	1	1	1	0	0	97.3	97.5
20	500	1	1.5	1	1	0	0	97.3	97.3
20	500	1.5	1	1	1	-2	0	97.5	97.2
20	500	1.5	1.5	1	1	-2	0	97.6	97.5

Comparison between univariate and multivariate meta-analysis in simulations

Comparison of point estimates

The point estimates for summary log odds ratios were very similar under univariate and multivariate meta-analysis. This is true for mutually exclusive outcomes (Figure 14), for outcomes with an is-subset-of relationship (Figure 15), and for sensitivity analysis scenarios where half of the studies did not report results for the second outcome (not shown). This finding is congruent with the conclusions of the empirical analyses.

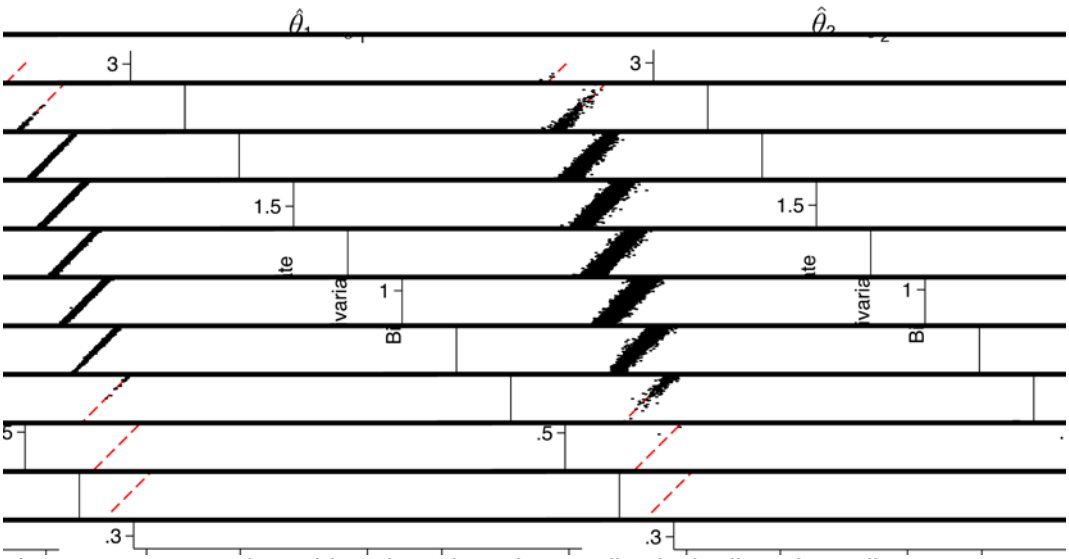
Figure 14. Summary odds ratios from univariate and bivariate meta-analysis across all simulations (mutually exclusive outcomes)



The axes correspond to odds ratios. The reference line is the line of equality.

In Figure 14 the scatter of the simulation points is identical for the two outcomes. This is because for mutually exclusive outcomes simulation results and conclusions are exactly symmetric with respect to the two outcomes. By contrast, for outcomes with an is-subset-of relationship (Figure 15), the random scatter of the points is greater for the first outcome than for the second outcome. This is expected, as the number of those experiencing the second outcome within each simulated study arm is larger, resulting in lower sampling variance.

Figure 15. Summary odds ratios from univariate and bivariate meta-analysis across all simulations (outcomes with an is-subset-of relationship)

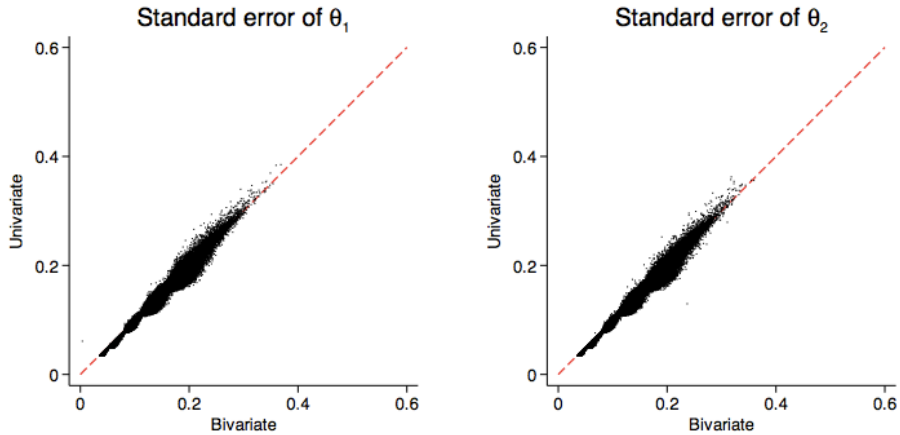


The axes correspond to odds ratios. The reference line is the line of equality.

Comparison of standard errors for the mean effects

As shown in Figure 16 for mutually exclusive outcomes and Figure 17 for outcomes with an is-subset-of relationship, the standard errors of the meta-analysis means are not substantially different in univariate versus bivariate analyses. This is congruent with the empirical results.

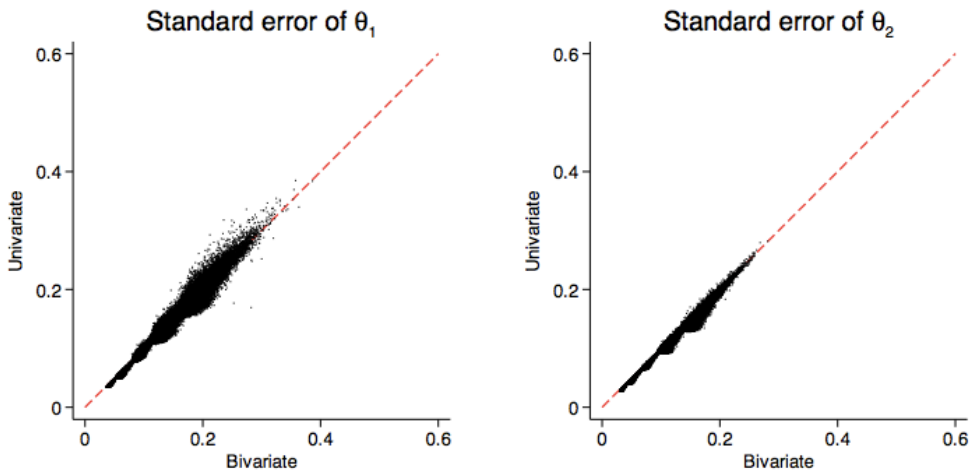
Figure 16. Standard errors of summary odds ratios with univariate and bivariate meta-analysis across all simulations (mutually exclusive outcomes)



The reference line is the line of equality.

Again, note that the two panels in Figure 16 are identical, because the simulation is symmetric with respect to the two mutually exclusive outcomes. In Figure 17, the standard errors for the second outcome are smaller than those for the first outcome; the explanation is the same as for Figure 15.

Figure 17. Standard errors of summary odds ratios from univariate and bivariate meta-analysis across all simulations (outcomes with an is-subset-of relationship)

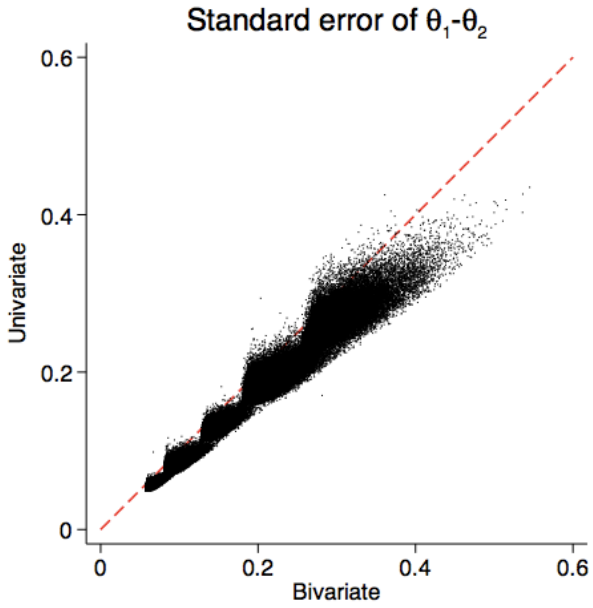


The reference line is the line of equality.

Comparison of standard errors of differences in the mean effects

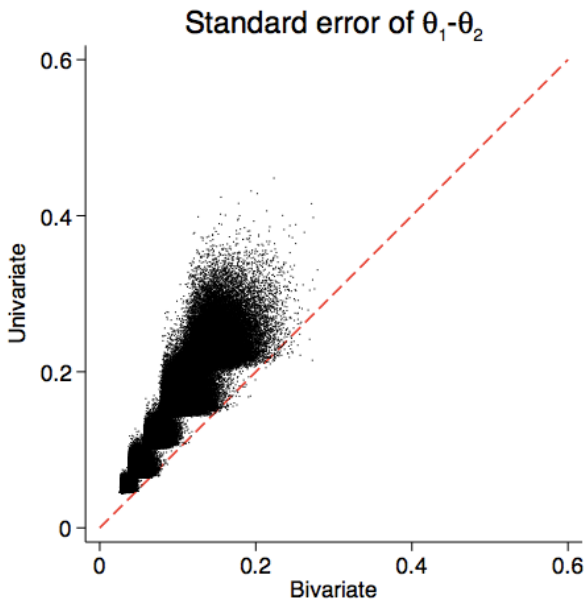
The most pronounced differences between univariate and bivariate meta-analyses pertain to the standard errors of differences in the summary estimates (or standard errors of linear combinations of the summary estimates in general[3,9]). In simulations of mutually exclusive outcomes, the between-study correlation is most often negative, and thus the standard error of the difference in the two means tends to be larger with bivariate compared to univariate meta-analysis (Figure 18). In simulations of outcomes that have an is-subset-of relationship the standard error with univariate meta-analyses tends to be larger than the standard error with multivariate meta-analyses (Figure 19), because the correlations between the outcomes were simulated to be positive, and are estimated as such.

Figure 18. Comparison of standard errors of the difference in the log summary odds ratios of two mutually exclusive outcomes with univariate and bivariate meta-analysis



The reference line is the line of equality.

Figure 19. Comparison of standard errors of the difference in the log summary odds ratios of two outcomes that have an is-subset-of relationship with univariate and bivariate meta-analysis



The reference line is the line of equality.

Discussion

We performed a large-scale empirical comparison of univariate and multivariate meta-analysis using data from the Cochrane Library of Systematic Reviews, and complemented it with a simulation study. Overall, univariate and multivariate methods yield numerically similar means and confidence intervals, suggesting that systematic review conclusions are not sensitive to this particular choice of methods. However, the confidence intervals of relative odds ratios between the pairs of outcomes can differ substantially between univariate and multivariate meta-analysis.

It appears that, as long as we focus on summaries for individual outcomes (and the respective confidence intervals) the choice between univariate and multivariate meta-analysis has limited practical importance. This is supported by our simulation analyses, and is congruent with the numerical results in the worked examples of several methodological papers introducing or reviewing methods for multivariate meta-analysis.[2,3,5,6,9,10,42-44] The main observations probably apply to all joint meta-analyses of two or more outcomes, irrespective of whether information on within-study correlations of treatment effects are exactly recoverable or have to be imputed, because the meta-analysis methodologies are the same.

It is not clear how often systematic reviewers face the methodological dilemma explored in this work. In our empirical evaluation, out of 1919 reviews with at least one binary meta-analysis, 29 (1.5%) reviews had at least one pair of meta-analyses that fulfilled our eligibility criteria. This proportion is probably an underestimate. We used outcomes exactly as defined by the Cochrane reviewers, and did not make any effort to redefine them to facilitate joint meta-analysis. Cochrane reviews include only univariate meta-analysis; if they were routinely performing multivariate analyses, they might have reviewed a larger number of outcomes. Further, we limited our analyses to examples where counts of combinations of outcomes are exactly recoverable from data used in univariate meta-analysis. However, it is possible that complete data on combinations of categorical outcomes can be obtained by contacting primary study authors, or even with care and perseverance during extraction of data from published articles (##redacted example as reference##).

So should one use separate or joint meta-analysis for sets of outcomes that can be approached with either method? In theory, the decision on performing separate versus joint meta-analyses depends on the underlying assumptions that the researcher is prepared to make about the data. Ideally, these decisions should be made early in the analysis, and not after examining the data. In fairness, however, and provided that one is not interested in linear combinations of treatment effects across outcomes (e.g., log relative odds ratios), decisions between univariate and joint meta-analysis seem to have theoretical rather than practical interest.

References

1. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *Stat Med*. Jan 26 2011.

2. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV, eds. *The handbook of research synthesis.* Vol 2nd. New York: Russel Sage Foundation; 1994:339-355.

3. Trikalinos TA, Olkin I. A method for the meta-analysis of mutually exclusive binary outcomes. *Stat Med*. 2008;27(21):4279-4300.

4. Arends LR. *Multivariate meta-analysis: Modelling the heterogeneity. Thesis dissertation*. 2006.

5. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Stat Med*. 1996;15(5):537-557.

6. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Stat Med*. 1998;17(22):2537-2550.

7. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat Med*. May 30 2010;29(12):1282-1297.

8. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Statistical methods in medical research*. Feb 16 2012.

9. Trikalinos TA, Olkin I. Meta-analysis of effect sizes reported at multiple time points: a multivariate approach. *Clin Trials*. 2012;XX(XX):XX-XX (in press).

10. Peter I, Crosier MD, Yoshida M, et al. Associations of APOE gene polymorphisms with bone mineral density and fracture risk: a meta-analysis. *Osteoporosis International*. Apr 2011;22(4):1199-1209.

11. Greenhalgh J, Hockenhull J, Rao N, Dundar Y, Dickson RC, Bagust A. Drug-eluting stents versus bare metal stents for angina or acute coronary syndromes. *Cochrane Database Syst Rev*. 2010(5):CD004587.

12. Rostom A, Dube C, Wells G, et al. Prevention of NSAID-induced gastroduodenal ulcers. *Cochrane Database Syst Rev*. 2002(4):CD002296.

13. Halliday HL, Ehrenkranz RA, Doyle LW. Late (>7 days) postnatal corticosteroids for chronic lung disease in preterm infants. *Cochrane Database Syst Rev*. 2009(1):CD001145.

14. Barrington KJ, Finer N. Inhaled nitric oxide for respiratory failure in preterm infants. *Cochrane Database Syst Rev*. 2010(12):CD000509.

15. Gong Y, Huang ZB, Christensen E, Gluud C. Ursodeoxycholic acid for primary biliary cirrhosis. *Cochrane Database Syst Rev*. 2008(3):CD000551.

16. Hodson EM, Willis NS, Craig JC. Corticosteroid therapy for nephrotic syndrome in children. *Cochrane Database Syst Rev*. 2007(4):CD001533.

17. Brion LP, Bell EF, Raghuveer TS. Vitamin E supplementation for prevention of morbidity and mortality in preterm infants. *Cochrane Database Syst Rev*. 2003(4):CD003665.

18. Brown J, O'Brien PM, Marjoribanks J, Wyatt K. Selective serotonin reuptake inhibitors for premenstrual syndrome. *Cochrane Database Syst Rev*. 2009(2):CD001396.

19. Al-Inany HG, Youssef MA, Aboulghar M, et al. Gonadotrophin-releasing hormone antagonists for assisted reproductive technology. *Cochrane Database Syst Rev*. 2011(5):CD001750.

20. Draper BH, Morroni C, Hoffman M, et al. Depot medroxyprogesterone versus norethisterone oenanthate for long-acting progestogenic contraception. *Cochrane Database Syst Rev*. 2006(3):CD005214.

21. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev*. 2006(3):CD004454.

22. Renfrew MJ, McCormick FM, Wade A, Quinn B, Dowswell T. Support for healthy breastfeeding mothers with healthy term babies. *Cochrane Database Syst Rev*. 2012;5:CD001141.

23. Hofmeyr GJ, Xu H. Amnioinfusion for meconium-stained liquor in labour. *Cochrane Database Syst Rev*. 2010(1):CD000014.

24. Soares-Weiser K, Goldberg E, Tamimi G, Pitan OC, Leibovici L. Rotavirus vaccine for preventing diarrhoea. *Cochrane Database Syst Rev*. 2004(1):CD002848.

25. Hodson EM, Craig JC, Strippoli GF, Webster AC. Antiviral medications for preventing cytomegalovirus disease in solid organ transplant recipients. *Cochrane Database Syst Rev*. 2008(2):CD003774.

26. Wilhelmus KR. Antiviral treatment and other therapeutic interventions for herpes simplex virus epithelial keratitis. *Cochrane Database Syst Rev.* 2010(12):CD002898.

27. Webster AC, Ruster LP, McGee R, et al. Interleukin 2 receptor antagonists for kidney transplant recipients. *Cochrane Database Syst Rev*. 2010(1):CD003897.

28. Webster A, Woodroffe RC, Taylor RS, Chapman JR, Craig JC. Tacrolimus versus cyclosporin as primary immunosuppression for kidney transplant recipients. *Cochrane Database Syst Rev*. 2005(4):CD003961.

29. Schieppati A, Perna A, Zamora J, Giuliano GA, Braun N, Remuzzi G. Immunosuppressive treatment for idiopathic membranous nephropathy in adults with nephrotic syndrome. *Cochrane Database Syst Rev*. 2004(4):CD004293.

30. Furukawa T, McGuire H, Barbui C. Low dosage tricyclic antidepressants for depression. *Cochrane Database Syst Rev*. 2003(3):CD003197.

31. Carlisle JB, Stevenson CA. Drugs for preventing postoperative nausea and vomiting. *Cochrane Database Syst Rev*. 2006(3):CD004125.

32. Bell RF, Dahl JB, Moore RA, Kalso E. Perioperative ketamine for acute postoperative pain. *Cochrane Database Syst Rev*. 2006(1):CD004603.

33. Stacey D, Bennett CL, Barry MJ, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2011(10):CD001431.

34. Oates-Whitehead RM, Baumer JH, Haines L, et al. Intravenous immunoglobulin for the treatment of Kawasaki disease in children. *Cochrane Database Syst Rev*. 2003(4):CD004000.

35. Hodnett ED, Gates S, Hofmeyr GJ, Sakala C, Weston J. Continuous support for women during childbirth. *Cochrane Database Syst Rev*. 2011(2):CD003766.

36. Duley L, Henderson-Smart DJ, Meher S, King JF. Antiplatelet agents for preventing pre-eclampsia and its complications. *Cochrane Database Syst Rev*. 2007(2):CD004659.

37. Hatem M, Sandall J, Devane D, Soltani H, Gates S. Midwife-led versus other models of care for childbearing women. *Cochrane Database Syst Rev*. 2008(4):CD004667.

38. Furukawa TA, Streiner DL, Young LT. Antidepressant and benzodiazepine for major depression. *Cochrane Database Syst Rev*. 2002(1):CD001026.

39. Amsallem E, Kasparian C, Haddour G, Boissel JP, Nony P. Phosphodiesterase III inhibitors for heart failure. *Cochrane Database Syst Rev.* 2005(1):CD002230.

40. Valgimigli M, Campo G, Arcozzi C, et al. Two-year clinical follow-up after sirolimus-eluting versus bare-metal stent implantation assisted by systematic glycoprotein IIb/IIIa inhibitor infusion in patients with myocardial infarction: results from the STRATEGY study. *J Am Coll Cardiol.* Jul 10 2007;50(2):138-145.

41. Valgimigli M, Percoco G, Malagutti P, et al. Tirofiban and sirolimus-eluting stent vs abciximab and bare-metal stent for acute myocardial infarction: a randomized trial. *JAMA.* May 4 2005;293(17):2109-2117.

42. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med.* Jan 15 2007;26(1):78-97.

43. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol.* 2007;7:3.

44. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med.* 2002;21(4):589-624.