Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2013;42:1120–1131
doi:10.1093/ije/dyt074


METHODOLOGY

# Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study

**Anna Chaimani,[1] Haris S Vasiliadis,[2,3] Nikolaos Pandis,[4,5] Christopher H Schmid,[6] Nicky J Welton[7] and Georgia Salanti[1]***

[1]Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece, [2]Department of Orthopaedics, School of Medicine, University of Ioannina, Ioannina, Greece, [3]Molecular Cell Biology and Regenerative Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [4]Department of Orthodontics and Dentofacial Orthopedics, Dental School, Medical Faculty, University of Bern, Bern, Switzerland, [5]Private practice, Corfu, Greece, [6]Center for Evidence Based Medicine and Department of Biostatistics, Program in Public Health, Brown University, Providence, RI, USA and [7]School of Social and Community Medicine, University of Bristol, Bristol, UK

*Corresponding author. Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, University Campus, Ioannina 45110, Greece. E-mail: gsalanti@cc.uoi.gr

**Accepted** 4 April 2013

**Background** Empirical research has illustrated an association between study size and relative treatment effects, but conclusions have been inconsistent about the association of study size with the risk of bias items. Small studies give generally imprecisely estimated treatment effects, and study variance can serve as a surrogate for study size.

**Methods** We conducted a network meta-epidemiological study analyzing 32 networks including 613 randomized controlled trials, and used Bayesian network meta-analysis and meta-regression models to evaluate the impact of trial characteristics and study variance on the results of network meta-analysis. We examined changes in relative effects and between-studies variation in network meta-regression models as a function of the variance of the observed effect size and indicators for the adequacy of each risk of bias item. Adjustment was performed both within and across networks, allowing for between-networks variability.

**Results** Imprecise studies with large variances tended to exaggerate the effects of the active or new intervention in the majority of networks, with a ratio of odds ratios of 1.83 (95% CI: 1.09,3.32). Inappropriate or unclear conduct of random sequence generation and allocation concealment, as well as lack of blinding of patients and outcome assessors, did not materially impact on the summary results. Imprecise studies also appeared to be more prone to inadequate conduct.

**Conclusions** Compared to more precise studies, studies with large variance may give substantially different answers that alter the results of network meta-analyses for dichotomous outcomes.

**Keywords** Multiple-treatments meta-analysis, indirect comparison, mixed-treatment comparison, small-study effects, publication bias

## Introduction

Empirical research has suggested that smaller studies tend to show larger treatment effects than do larger studies.[1–3] This phenomenon, known as the small-study effect, can be partly explained by publication bias; small studies with non-significant treatment effects are less likely to be published. Other explanations for small-study effects, such as the association between small study size and a lower study quality that leads to overestimation of treatment effects, very likely operate alone or in combination with publication bias. In the case of randomized controlled trials (RCTs), allocation concealment and blinding have been found to be important quality-related trial characteristics that can influence the results of individual studies or meta-analyses.[4–6] This association was recently confirmed by the largest such study to date,[7] although the strength of the association was found to be less than in previous analyses.

Because both study size and variability in the treatment effects among the individual patients within a study may be associated with the magnitude of effect estimates, statistical models exploring the effects of these study characteristics usually use the precision of the study summary estimates (or a function of the precision, such as the standard error or variance) as an explanatory factor. Several regression-based methods have been proposed to model the relationship between effect size and some measure of its precision or other study characteristics.[1,8,9] These methods require a sufficient number of available studies, and over a range of study characteristics, to estimate the impact of study characteristics (represented by a bias parameter) on the results of meta-analysis.[10] For this reason, meta-epidemiological approaches, which consider collections of independent meta-analyses and assume that the bias across these meta-analyses is of comparable magnitude and direction, are employed to increase the number of studies in the analysis and improve the power of the analysis.[5,10–12] However, the comparability of the bias parameters across meta-analyses from different fields of clinical research is often questionable.

Network meta-analysis is an extension of conventional (pairwise) meta-analysis that synthesizes evidence from systematic reviews that compare multiple treatments.[13,14] Network meta-analysis data have not yet been considered as a means for evaluating bias in meta-epidemiological studies, although these data arguably provide a larger evidence base upon which to address such methodological questions. First, systematic reviews aiming to compare multiple treatments produce a network of comparisons for the same outcome and population, and therefore provide a natural pool of meta-epidemiological data for which it could be assumed that bias is acting in a similar way across trials that have similar characteristics. Second, network meta-analysis is more powerful than pairwise meta-analysis when adjustment is performed for potentially biased effect sizes.[15] Third, the use of network meta-analysis in meta-epidemiology (which we shall refer to as network meta-epidemiology) to estimate bias parameters exploits the assumption that bias parameters are more similar across comparisons within the same treatment network than they are across networks. The use of network meta-epidemiology has been exemplified by Salanti *et al*.[16] The aim of the present paper is to use network meta-epidemiology to determine the association of trial characteristics (generation of allocation sequence, allocation concealment, blinding) with treatment effects estimated in network meta-analysis and their connection with study precision. For this purpose, we use a collection of published networks of treatments.

## Methods

### Selection of networks of interventions

We searched in PubMed for meta-analyses of RCTs published by the end of March 2011 in which at least three treatments were included and the data had been analysed with a valid statistical method for indirect comparisons or network meta-analysis (also called multiple-treatments meta-analysis or mixed-treatment comparison). We applied no restriction regarding the type of outcome measure for the individual studies (dichotomous, continuous, etc.), nor for the effect sizes of meta-analyses (odds ratio (OR), risk ratio (RR), mean difference (MD), etc.). To evaluate the impact of the trial characteristics and imprecise-study effects, we included only networks in which all treatments were compared to a common comparator (star-shaped networks). Such networks do not have closed loops. The advantage of using star-shaped networks is that the direction of possible bias is expected to operate against the common comparator, typically an inactive or older standard treatment in each network's field. We excluded diagnostic test accuracy studies and networks in which the number of treatments was greater than the number of trials (networks with sparse information). To further investigate the presence of imprecise-study effects, we enriched the database with full networks (which include closed loops) of at least four treatments that involved an obvious reference intervention. Details of the search code can be found in the Supplementary Appendix, available as Supplementary data at *IJE* online.

### Data extraction

From every published paper included in our study, we extracted the method used for the indirect comparison, as well as the number of studies and interventions and data for the primary outcome (or, if this was not specified, the outcome presented first in the

meta-analysis). If outcome data or risk of bias information were not available, we contacted the authors. Arm-level data were preferred to study-level data[17] when both were reported; a mixture of both was used if some studies reported trial-level and some reported arm-level data.[18] Two authors independently assessed the risk of bias in all RCTs of all included networks for which risk of bias data were not available either in the publication or from the authors upon request. Based on the Cochrane Risk of Bias Tool,[19] we sought information regarding the random sequence generation, allocation concealment, and blinding of patients and outcome assessors. For each item, studies were classified into the three groups; being at low, unclear, or high risk of bias.

## Statistical analysis

We analysed all included star-shaped networks using standard Bayesian network meta-analysis models.[14,15] These models assume consistency in the estimated effect sizes. This means that if $\mu_{z,c}$ is the effect of treatment $c$ vs. the reference treatment (the common comparator) in star network $z$, and $\mu_{z,j}$ is the effect of $j$ vs. the reference, then the effect of $c$ vs. $j$ is $\mu_{z,cj} = \mu_{z,c} - \mu_{z,j}$. We used a random effects model that allows for between-study variation in the estimated effect sizes and a common heterogeneity variance ($\tau_z^2$) for all comparisons in a network $z$.

To account for the impact of trial characteristics and imprecise-study effects, we employed network meta-regression models.[20,21] We used an indicator variable to represent each bias item (categorized as low risk vs. high or unclear risk) and the variance of the observed effect size[22] to represent an imprecise-study effect (studies with treatment effects with large variances). Adjustment can take place both on models for study-level or arm-level data (a detailed description is given in the Supplementary Appendix).

To ensure enough power to perform meta-regression, we initially assessed the effects of each risk of bias item and study precision in separate meta-regression models within each star-shaped network (indexed with $z$) that included at least 10 studies.[23] We re-expressed each covariate so that for both harmful and beneficial outcomes a positive regression coefficient, $\beta$, indicated that less precise studies or studies at unclear or high risk of bias overestimated the effectiveness or safety of each treatment compared to the reference treatment. We used the enriched database (including the full networks) to perform sensitivity analysis for the investigation of imprecise-study effects and we estimated the regression coefficient for each full network assuming imprecise head-to-head trials have no impact on the results.

We first assumed that the impact of trial characteristics was identical for each comparison of a treatment $k$ with the reference treatment within the same network (i.e. that $\beta_{z,k} = B_z$ (a single fixed-effect coefficient)). As sensitivity analyses, we assumed that regression coefficients $\beta_{z,k}$ were different yet exchangeable across comparisons within the same network, by sharing a normal distribution with a common mean, $B_z$, and variance $\sigma_z^2$, with the latter expressing the variability between the coefficients in the same network.

In the network meta-epidemiological model, we linked the network-specific coefficients across networks into a 'joint model.'[16] This allowed networks with few studies to borrow strength from those with more studies, and increased the power of the analysis while yielding an estimated overall coefficient, $B_{overall}$. We allowed for between-networks variability by assuming $B_z \sim N(B_{overall}, \omega^2)$, and in a sensitivity analysis we fitted a common (fixed) coefficient across networks ($B_z = B_{overall}$). The network meta-epidemiological model linked only networks estimating outcomes using the same effect size measure [for dichotomous outcomes the OR, for continuous outcomes the MD and for survival the hazard ratio (HR)]; networks with different effect size measures were excluded from this analysis. For dichotomous outcomes, the association is presented as the ratio of ORs (ROR), which is the exponent of $B_z$ or $B_{overall}$.

A graphical representation of the network meta-epidemiological model can be found in the Supplementary Appendix (Supplementary Figure 1). All models are presented in detail in the Appendix.

## Sensitivity and subgroup analyses

Additional analyses were conducted to check the sensitivity of the results to the measure of precision of the observed effect size used as an explanatory factor (variance, standard error, inverse of variance, or square root of inverse variance) in the analysis of imprecise-study effects. We performed two additional analyses in which we included all networks with dichotomous outcomes (star and full networks), to evaluate the impact of studies with small sample size. These were defined as studies with a total sample size of less than 200 (as suggested in Zhang *et al.*[24]) or less than 300 (based on the present data; see Supplementary Appendix for details) participants. We assessed the robustness of results to the exclusion of networks with very large or small coefficients compared to the other networks. Such networks may violate the exchangeability assumption underlying the coefficients across networks. We performed a sensitivity analysis for the network meta-epidemiological model using the RR instead of the OR. Evaluation of the impact of risk of bias items was done separately for mortality and non-mortality outcomes.[5] We extracted risk of bias information only for a subset of the networks (relying on previously reported risk of bias classifications in the remainder of cases, for which it is possible that different criteria were used). We therefore performed a sensitivity analysis that included only networks for which we extracted data according to our criteria as described in the

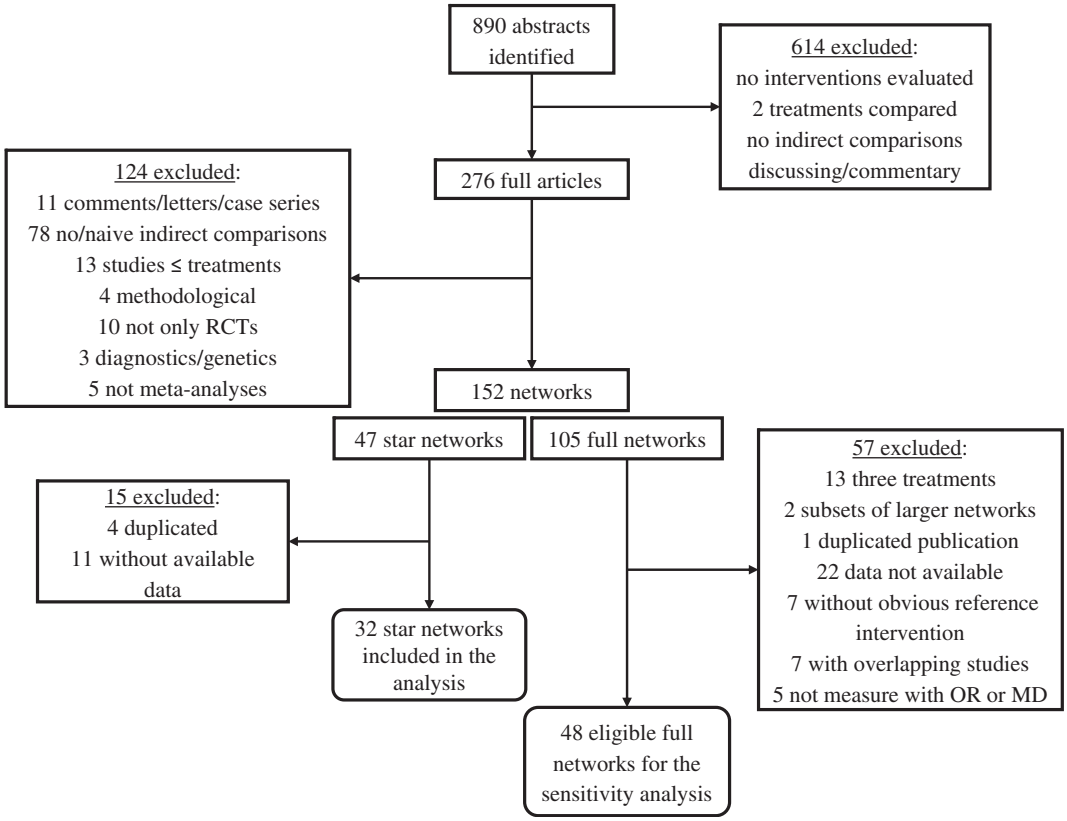


**Figure 1** Flow chart of the study selection

Supplementary Appendix. To check whether different types of interventions affected the results of the adjustment for risk of bias items, we excluded networks with non-pharmacological interventions from our analysis.

### Model selection and implementation

We evaluated the parsimony of all models according to the deviance information criterion (DIC), lower values of which suggest a better compromise between model fit and model complexity.[25] All models were fitted using Markov chain Monte Carlo simulation in the freely available software WinBUGS 1.4.3.[26] Normal vague priors were given to location parameters. We assumed a half-normal prior distribution $\tau \sim N(0,1)$ with $\tau \geq 0$ for the between-study heterogeneity standard deviation, $\tau$, and a uniform (0,3) distribution for the standard deviation, $\omega$, of the imprecise-study effects coefficients. All results are presented as posterior medians with the 95% credible interval. More technical details and alternative prior distributions for $\omega$ are presented in the Supplementary Appendix.

## Results

### Characteristics of included networks of interventions

The search identified 890 relevant abstracts, of which 276 were assessed as potentially eligible and whose full articles were screened. After the classification of studies into networks with and without closed loops, we ended up with 32 star-shaped networks (613 studies) suitable for inclusion (see Figure 1 for details of the search results and selection process). All of the networks were published after 1999 and cover a variety of medical fields, as shown in Supplementary Table 1 (available as Supplementary data at *IJE* online). The method of indirect comparison used were Bucher's method (known also as the method of adjusted indirect comparison) (50%), meta-regression approaches (25%), and Bayesian network meta-analysis (25%). Twenty-two (69%) of the star-shaped networks included 10 or more studies, and the number of treatments compared ranged from 3 to 20. Data about risk of bias in the included studies were obtained for these 32 star-shaped networks (available in the original publication, extracted, or provided by the authors). The number of studies determined to be at low, unclear, and high risk of bias for each network are presented in Supplementary Table 2 (available as Supplementary data at *IJE* online). We included 22 star-shaped networks with 10 or more studies (a total of 545 studies) for the independent network-specific adjustment, in which each network was analysed separately estimating independent network-specific coefficients. Adjustment of each network using exchangeable coefficients across treatment comparisons yielded

similar results and similar model fit to those obtained with fixed coefficients. We therefore report the latter as the simpler model, with more precise estimates. Twenty star-shaped networks (after the exclusion of networks with overlapping studies) (358 studies) reported dichotomous outcomes measured with OR, and were included in the meta-epidemiological model. The remaining 5 star-shaped networks with continuous data, 4 networks reporting HR (for each included study), 1 network with rate data as well as 2 networks that used RR as effect size measure, were not synthesized because of the small numbers of networks. We further identified 34 full networks (934 studies) with dichotomous outcomes measured with OR, and 13 full networks (358 studies) with continuous outcomes measured with MD, which we included in the meta-epidemiological model for imprecise-study effects. These were combined (after excluding overlapping networks) with 18 star-shaped networks with dichotomous data and with 2 star networks with continuous data, respectively. Only 7 networks (4 star and 3 full networks) reporting HR were identified and were therefore not synthesized (see Supplementary Table 3 for the characteristics of full networks).

## Adjustment for risk of bias items

Network-specific effects for the four risk of bias items did not show any impact on the estimated treatment effects. Of the 22 networks with 10 or more studies, 60%, 55%, 56%, and 61% resulted in positive coefficients for allocation concealment, sequence generation, blinding of patients, and outcome assessors, respectively (see Table 1). None of the coefficients was statistically significant or close to significance. The heterogeneity also did not change substantially, with differences being smaller than 3.5% of the estimated heterogeneity before the adjustment.

Figure 2 shows the network meta-epidemiological summary effects from the 20 star-shaped networks when exchangeable coefficients were assumed across networks and fixed coefficients within networks. No evidence is provided for the impact of studies with high or unclear risk of bias on the relative effectiveness or safety of the interventions. The between-networks standard deviation ($\omega$) (on ln($OR$) scale) of the summary ROR was 0.91 (0.75, 1.09), 0.98 (0.83, 1.18), 1.16 (0.95, 1.43), and 1.15 (0.83, 1.60) for random sequence generation, allocation concealment, and blinding of patients and outcome assessors, respectively, and did not change when different prior distributions were used. The meta-epidemiological models that did not account for between-networks variability (estimating a fixed coefficient across networks) led to similar conclusions. The DIC values were similar before and after the adjustment in almost all analyses, and changes were smaller than the three-point threshold that can support differences in model parsimony (Supplementary Table 4, available as Supplementary data at *IJE* online).[25] Subgrouping the networks according to the type of outcome (mortality vs. other) did not change the results (Figure 2). We also compared the rankings of treatments within each network as estimated by the meta-epidemiological and the unadjusted models. Adjustment for blinding of patients yielded small differences in the hierarchy of the treatments in three networks; blinding of outcome assessors and sequence generation affected two networks; and allocation concealment changed one network. When we analysed only networks for which we had assessed the risk of bias in individual studies (12 networks), the RORs for random sequence generation, allocation concealment, and blinding of patients and outcome assessors were 0.83 (0.63, 1.12), 0.97 (0.74, 1.30), 1.07 (0.80, 1.46) and 1.22 (0.83, 1.88), respectively. Excluding networks with non-pharmacological interventions (2 networks) yielded the same results as did the primary analysis.

## Adjustment for study precision

Of 22 networks including 10 or more studies, 18 (81.8%) resulted in positive imprecise-study effects coefficients, implying that imprecise studies exaggerated the treatment effects compared to the control intervention in most of the networks. Eight of the positive coefficients (36.4%) and 1 of the 4 negative coefficients showed a large effect of imprecise studies in the respective networks (accounting for the different effect size measures between networks), and were also statistically significant (see Table 1). In 4 of the 8 networks with positive significant coefficients, the DIC of the adjusted model was considerably lower than that of the unadjusted model, showing an important improvement in model parsimony after the adjustment (Supplementary Table 3, available as Supplementary data at *IJE* online). Heterogeneity decreased in 13 (40.6%) networks after the adjustment for imprecise-study effects, with a relative drop ranging from 7.1% to 39.5%; 13 networks showed a relative increase in heterogeneity, of between 1.4% and 9.1% (Figure 3a). Figure 3b shows the corresponding reduction in the comparison-specific effect sizes, indicating that in most of the networks, the relative effects of active or newer treatments versus the common comparator were exaggerated in less precise studies. The meta-epidemiological summary (ROR) estimate accounting for between-studies variability in ORs was 1.84 (1.09, 3.32), with between-networks standard deviation, $\omega = 0.83$ (0.41, 1.48) (Figure 4). The model that ignored the between-networks variability (fixed coefficient) further increased the precision of the analysis and resulted in an overall common imprecise-study effects ROR equal to 1.38 (1.11, 1.70). The sum of DICs from the unadjusted models for all star-shaped networks was 1291, which was considerably higher than the DIC of the joint model

**Table 1** Network-specific coefficients for all networks with at least 10 studies, as derived from the separate analysis[a,b]

| Network identification number | Imprecise-study effects | Sequence generation | Allocation concealment | Blinding of patients | Blinding of outcome assessors | Measure of effect |
|---|---|---|---|---|---|---|
| 1 | 9.87 (2.66, 49.90) | 12.68 (0.52, 237.46) | – | 2.14 (0.16, 24.29) | 2.14 (0.16, 24.29) | Ratio of odds ratios |
| 3 | 0.40 (0.17, 0.97) | 0.88 (0.37, 1.99) | 1.17 (0.46, 2.86) | 1.39 (0.33, 6.11) | 1.36 (0.33, 5.64) | |
| 5 | 2.10 (1.08, 4.44) | 1.62 (0.57, 5.37) | 1.73 (0.55, 5.53) | 1.40 (0.47, 4.06) | – | |
| 11 | 2.48 (0.24, 27.66) | 1.03 (0.62, 1.68) | – | 1.77 (0.84, 3.63) | 1.75 (0.67, 4.90) | |
| 12 | 1.26 (0.66, 2.61) | 0.73 (0.49, 1.07) | 1.15 (0.66, 1.99) | 1.01 (0.62, 1.55) | 1.08 (0.68, 1.65) | |
| 13 | 4.71 (1.48, 17.81) | 0.41 (0.11, 1.65) | 1.36 (0.33, 6.49) | 1.00 (0.23, 4.57) | 2.18 (0.54, 9.58) | |
| 15 | 1.35 (0.79, 2.41) | 1.04 (0.53, 1.62) | 0.95 (0.53, 1.77) | 0.95 (0.58, 1.95) | – | |
| 17 | 0.99 (0.28, 3.53) | 1.11 (0.36, 3.25) | 0.92 (0.41, 2.08) | – | – | |
| 18 | 1.07 (0.80, 1.43) | 1.06 (0.81, 1.32) | 0.96 (0.77, 1.19) | 1.27 (0.95, 1.82) | – | |
| 26 | 0.57 (0.06, 4.85) | 0.61 (0.23, 1.65) | 1.03 (0.33, 3.22) | – | 1.31 (0.53, 3.35) | |
| 27 | 6.96 (1.88, 27.39) | 1.68 (0.89, 3.25) | 2.10 (0.95, 4.66) | 1.65 (0.82, 3.35) | 1.65 (0.75, 3.71) | |
| 30 | 4.66 (1.34, 22.20) | 0.84 (0.13, 4.85) | 1.28 (0.23, 9.03) | 0.92 (0.31, 3.32) | 0.51 (0.18, 1.26) | |
| 31 | 0.57 (0.16, 1.88) | 0.54 (0.31, 0.94) | 0.61 (0.31, 1.13) | 0.57 (0.24, 1.27) | 0.57 (0.24, 1.28) | |
| 32 | 1.07 (0.50, 2.56) | 0.82 (0.55, 1.25) | 0.82 (0.55, 1.22) | 0.97 (0.61, 1.60) | – | |
| 21 | 6.96 (1.04, 47.47) | 0.76 (0.19, 3.13) | 1.06 (0.57, 2.08) | 1.14 (0.64, 1.92) | 1.21 (0.68, 2.01) | Ratio of risk ratios |
| 25 | 2.10 (0.97, 4.44) | 1.34 (0.94, 1.92) | 0.86 (0.58, 1.26) | 0.97 (0.68, 1.38) | 0.97 (0.68, 1.38) | |
| 28 | 1.04 (0.84, 1.30) | 0.94 (0.10, 9.30) | 0.86 (0.05, 15.03) | 0.66 (0.02, 17.12) | 1.09 (0.04, 27.11) | Ratio of rate ratios |
| 9 | 4.06 (1.28, 12.43) | 1.09 (0.81, 1.51) | 1.09 (0.90, 1.31) | – | – | Ratio of hazard ratios |
| 29 | 2.44 (0.00, 110.95) | 0.92 (0.68, 1.26) | 1.01 (0.76, 1.36) | – | – | |
| 19 | 4.56 (0.63, 8.58) | 0.02 (-0.26, 0.32) | – | –0.03 (–0.15, 0.10) | –0.01 (-0.14, 0.11) | Difference of standardized mean differences |
| 2 | 0.48 (–0.86, 1.84) | 0.73 (–2.01, 3.47) | 0.72 (–2.02, 3.49) | 0.19 (–1.78, 2.16) | 0.19 (–1.78, 2.16) | Difference of mean differences |
| 20 | 0.05 (–0.42, 0.49) | 0.07 (–0.97, 1.16) | 0.41 (–0.65, 1.57) | 0.68 (–0.38, 1.82) | 0.68 (–0.38, 1.82) | |

[a]Missing values are for networks in which all studies are at low or unclear/high risk of bias.
[b]Coefficients larger than 1 (for dichotomous data) or positive (for continuous data) indicate that less precise studies or studies at unclear/high risk of bias overestimate the effect of each treatment compared with the reference.

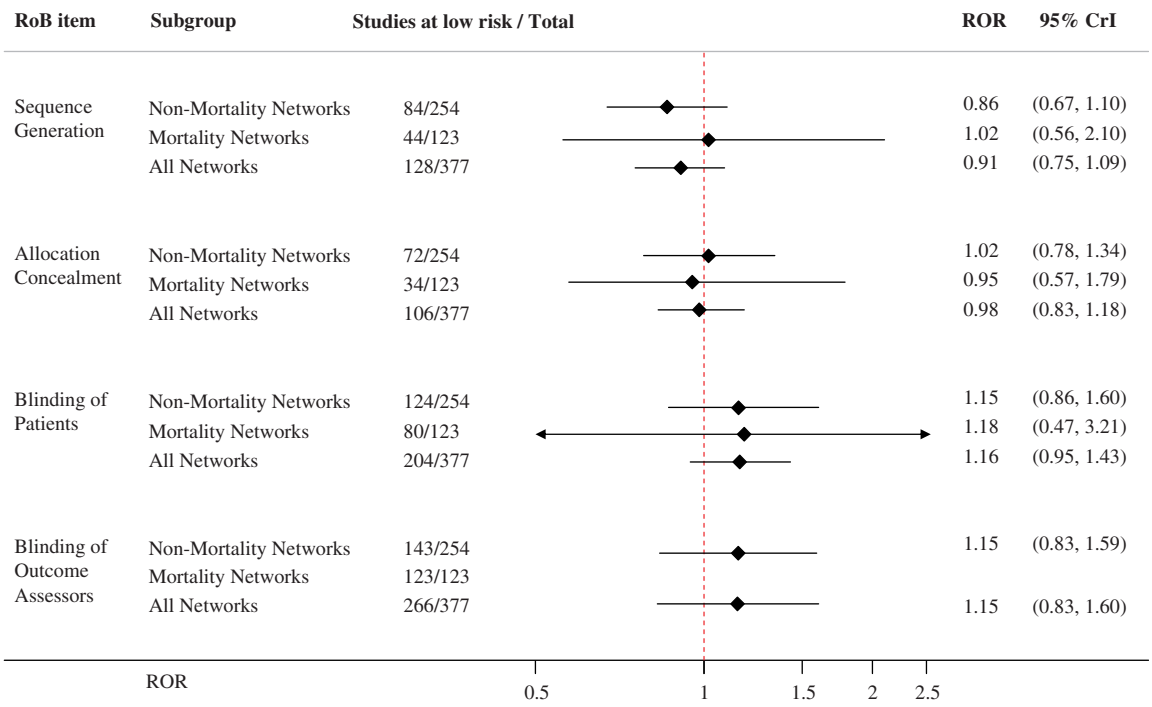


**Figure 2** Overall ratios of odds ratios (RORs) for each risk of bias component, derived from the joint analysis and for subgroups on mortality and non-mortality networks. Outcome assessors were defined as being blinded for all mortality networks. An ROR larger than 1 indicates that the effect of new or active treatments is exaggerated relative to the control intervention in studies at high or unclear risk of bias (CrI = credible interval)

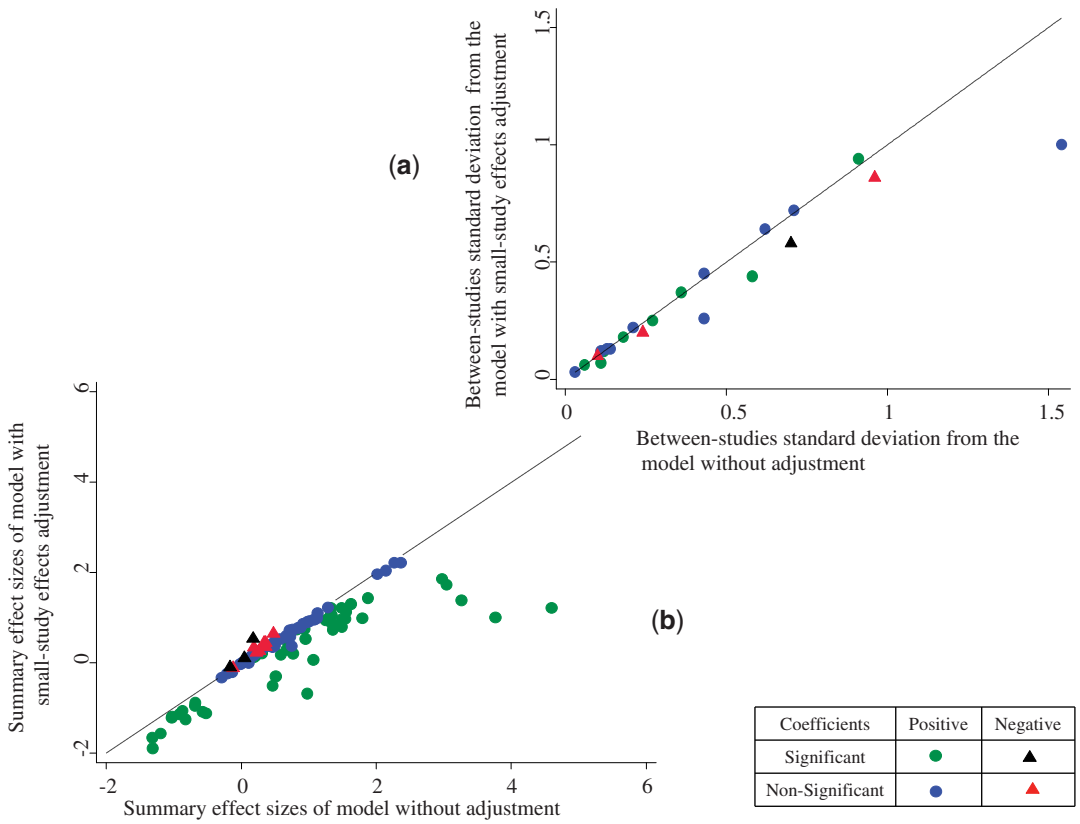


**Figure 3** Comparison of (a) between-study standard deviations, and (b) the pooled relative effect sizes of all treatments vs. the common comparator for all networks with at least 10 studies between the unadjusted model and the model adjusted for imprecise-study effects. The diagonal lines indicate equality of effects with and without adjustment

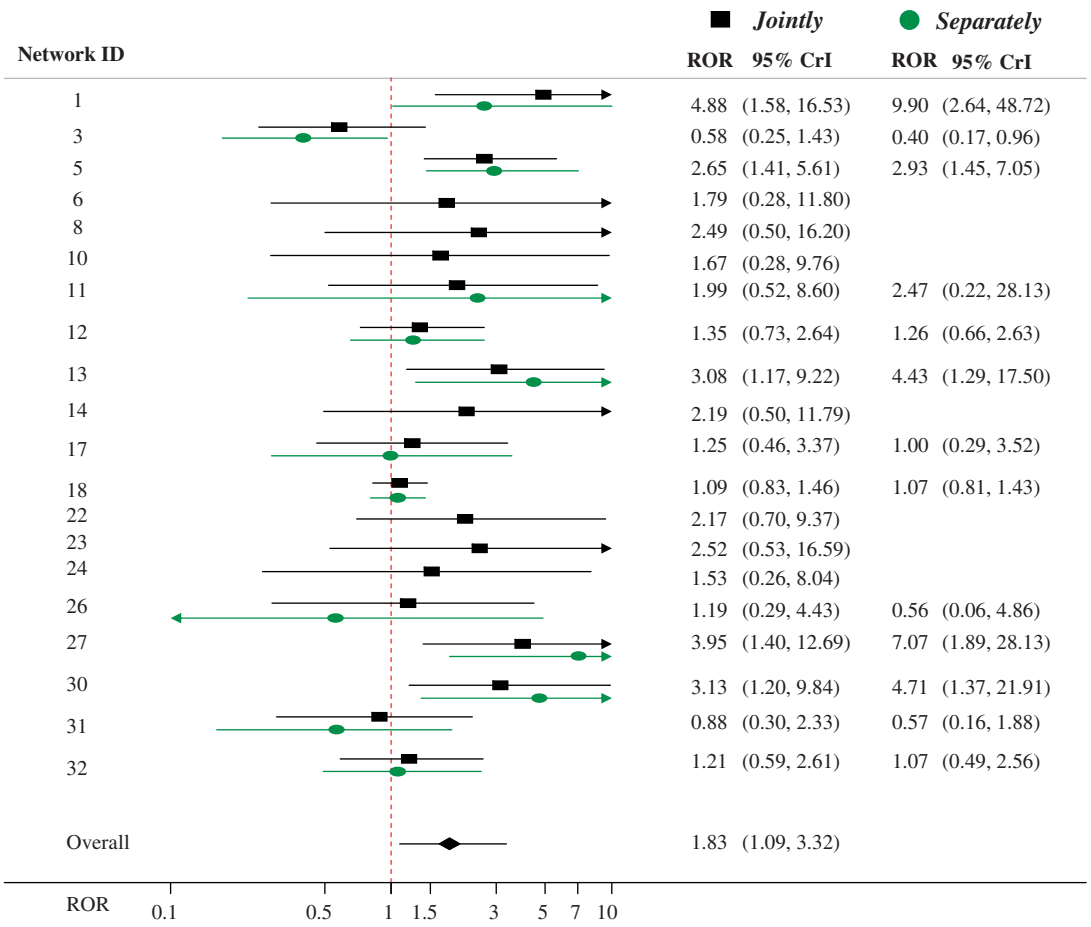


**Figure 4** Network-specific coefficients of imprecise-study effects for all networks with dichotomous outcome data, based on models for the separate analysis of each network and the joint analysis of all networks. A ratio of odds ratios (ROR) larger than 1 indicates that imprecise studies exaggerate the effect of new or active treatments relative to the control intervention. (CrI = credible interval)

(DIC = 1276), showing that adjustment led to more parsimonious models. Comparison of treatment ranks as estimated from the joint model and the unadjusted model for each network showed differences in treatment classification in three networks. More specifically, the common comparator interventions of these networks had a higher rank in the treatment hierarchy after the adjustment.

## Sensitivity analyses

Changing the measure of precision used as covariate for the adjustment of imprecise-study effects resulted in slightly different but compatible conclusions. The coefficient for the regression on the standard error was similar but less precise than in the primary analysis [ROR = 2.01 (1.02, 12.22)]. Using the inverse of variance or the square root of inverse variance as covariates gave estimates close to the null value of 1; however, the model fit and parsimony as conveyed by the DIC values was much worse for these two models (Supplementary Table 5, available as Supplementary data at *IJE* online).

Network 1[27] resulted in a considerably larger coefficient compared to those of the other networks, and might thus violate the exchangeability assumption (Figure 4). When, excluded, the meta-epidemiological summary ROR was estimated as 1.62 (1.00, 2.86). When we performed the analysis (with all star networks) on the log-risk ratio scale, the overall coefficient was similar to that of the primary analysis. Lastly, the inclusion of the 34 full networks in the analysis did not substantially change the magnitude of the summary coefficient, but did increase the precision of the estimate [ROR = 1.88 (1.43, 2.51)] (Supplementary Table 5, available as Supplementary data at *IJE* online). Using sample size thresholds of 200 and 300 patients, small trials appeared to exaggerate the ORs for the active or experimental interventions by on average 26% and 13%, respectively [RORs 1.26 (1.10, 1.46) and 1.13 (1.00, 1.28)]. When we fitted the meta-epidemiological model for the 15 networks (2 star and 13 full networks) using MD as the measure of effect, the coefficient estimate (difference of mean differences) was close to zero [$B = 0.02$ (–0.34, 2.38)]. The use of alternative

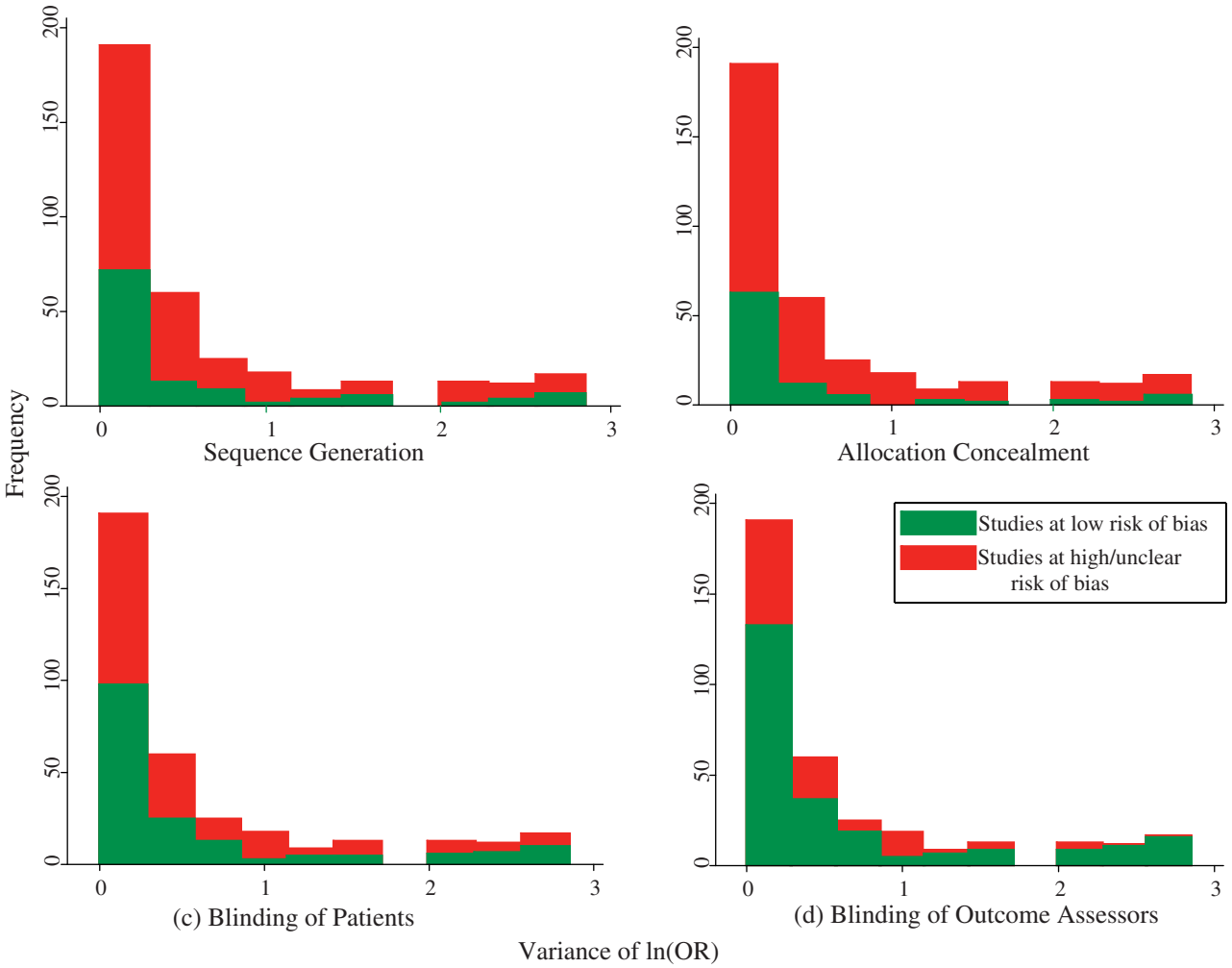


**Figure 5** Histograms showing the distributions of variance (of ln(*OR*)) for individual studies classified according to each risk of bias category for random sequence generation (a), allocation concealment (b), and blinding of patients (c) and outcome assessors (d)

distributions for the between-networks standard deviation, $\omega$, of the coefficients did not affect the results of any of the analyses.

### Association between study size and risk of bias components

Figure 5 shows the distribution of study variance for studies at low risk versus high or unclear risk of bias. Most studies with small variances were assessed as being at high or unclear risk of bias for sequence generation and allocation concealment. The opposite was observed for blinding of patients and outcome assessors, where more precise studies were more often adequately blinded.

## Discussion

To our knowledge, this is the first network meta-epidemiological study assessing the impact of imprecise-study effects and of four risk of bias items on treatment effects. Previous empirical studies of pairwise meta-analyses investigating the association between risk of bias items and treatment effects have yielded inconsistent conclusions. Balk *et al.*[28] found that poor randomization, allocation concealment, and lack of blinding did not affect the estimated treatment effects, which agrees with our findings. However, other meta-epidemiological studies[4,29] have shown that studies with inadequate conduct of these three risk of bias items tended to exaggerate the treatment effects. Siersma *et al.*[30] found that random generation of sequence allocation was an important predictor of trial results. In contrast to our study, as well as that of Balk *et al.*,[28] the results presented by Wood *et al.*[5] imply that subjective outcomes are affected by inadequate allocation concealment and lack of blinding of patients more than all-cause mortality and other objective outcomes. Also, Hrobjartsson *et al.*[31] found that blinding of outcome

assessors affected the treatment effect estimates in trials with binary subjective outcomes.

A recent combined analysis of seven meta-epidemiological studies[7] resulted in marginally exaggerated intervention effects in trials with inadequate or unclear conduct of random sequence generation, allocation concealment, and double-blinding. Our study's finding of weaker associations might be explained by limited power; our set of network meta-analyses included on average 31% fewer trials compared to Savovic *et al*.[7] Moreover, because network meta-analysis is a novel and resource-demanding methodology, investigators undertake it primarily to answer questions for which they know many valid studies exist. Additionally, the lack of association between risk of bias items and treatment effects might be due to random misclassification resulting from the different approaches and criteria used by the original authors of each network to assess the risk of bias in trials. However, in the networks for which we ourselves extracted risk of bias data, the RORs for the four risk of bias items did not differ from the RORs for which we did not extract such data. Our data do not highlight a clear association between study size and study quality, which was present in the analyses of Kjaergard *et al*.[32] and Nuesch *et al*.[2]

In our analysis, we considered trials at unclear and high risk of bias as having a similar effect. This assumption can be relaxed in models that attribute a probability of being at high risk to studies of unclear risk.[13] However, in our dataset, the number of studies at high risk of bias was very small, and such models do not have enough power to estimate effects. When risk of bias characteristics are unavailable, study precision could be considered as a proxy for them, but interpretation of the results in this case needs careful consideration. Furthermore, it is possible that other types of biases operate, such as selective outcome reporting and attrition bias, for which we did not account.

Our analyses illustrated an important impact of study precision on the results of network meta-analysis with dichotomous outcomes. If some comparisons in the network are informed by less precise estimates than others, inconsistency can occur.[33] Consequently, investigators should routinely evaluate the distribution of variance across comparisons included in a network, and should consider study precision as a possible source of inconsistency. In the present study we included only star-shaped networks, which do not allow statistical evaluation of the assumption of consistency.

Adjustment for imprecise studies can be done routinely through meta-regression. The models presented here rely on a range of study variances on which the coefficients are estimated. Estimation of regression coefficients is challenging when comparisons include studies of similar precision, especially in models with independent coefficients. The estimated intercept can be seen as the relative treatment effect when variance approximates zero. Because zero is outside the range of observed variances, it might be more interpretable to center the precision variable at the smallest observed variance. In some circumstances researchers may need to extrapolate beyond the range of observed precisions (e.g. if all the studies are very imprecise); caution is needed in doing so. An association between effect size and precision can be heuristically understood as small studies providing different results from larger studies, although sample size is only one of the components influencing the variance of the observed effect size. It is not possible to define what constitutes an imprecise or 'small' study unless one restricts its definition to a specific application and outcome (such as in the study by Nuesch *et al*.[2]).

Continuous measures are often used for subjective and secondary outcomes, and should therefore be more prone to selective outcome reporting resulting in a pronounced association between variance and study effect. However, we did not observe important imprecise-study effects in networks with continuous outcomes. This might be due to low power (only 16 networks with continuous data were identified) or to the larger heterogeneity in continuous effect measures.

In our analysis, we used the variance of the observed effect size as an explanatory variable for the between-studies variation of the treatment effect. A limitation of this approach for dichotomous outcomes is that because the observed log-odds ratio and its variance are both functions of the true log-odds ratio, an inherent correlation can be induced. However, use of the exact binomial likelihood, as we were able to do whenever arm-level data were available, mitigates this dependence.[33,34]

## Conclusions

In this network meta-epidemiological study of 32 star-shaped networks comprising 613 trials from several medical fields, we found evidence that imprecise studies produce larger effects than do more precise studies. However, we found no evidence supporting an association of effect size with other previously identified indicators of bias, such as random sequence generation, allocation concealment, and blinding. Subgroup-specific coefficients for specific types of outcomes estimated in our study could be used to form informative priors and adjust treatment effects in future network meta-analysis studies.[13]

As with conventional pairwise meta-analysis, network meta-analysis can be affected by the tendency of imprecise studies to overestimate the effectiveness or safety of the treatments, and this should be taken into account when conducting such an analysis. In this case, adjusted results from network meta-regression models should be presented extrapolated to the most precise study in the network.[22] Modified funnel

plots can also be used for this purpose,[34] and selection models can be fit in case the association between study size and effect is due to publication or reporting bias.[35]

## Supplementary Data

Supplementary Data are available at *IJE* online.

## Funding

This work was supported by the European Research Council (IMMA 260559 project, to A.C. and G.S.) and the US Agency for Healthcare Research and Quality (grant R01HS018574, to C.S.).

## Acknowledgements

We thank Drs Hofmeyr, Jansen, Loke, Mills, Nelson, Maison, Tramer, Piccini, Tudor-Smith, and Vandermeer for having provided us with data; Drs Avendano, Farietta, and Kong Win Chang for having provided risk of bias data, and Dr Trikalinos for his contribution to our search strategy.

**Conflict of interest:** None declared.

## References

[1] Moreno SG, Sutton AJ, Ades AE *et al*. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med Res Methodol* 2009;**9:**2.

[2] Nuesch E, Trelle S, Reichenbach S *et al*. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;**341:**c3515.

[3] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;**7:**1–76.

[4] Moher D, Pham B, Jones A *et al*. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352:** 609–13.

[5] Wood L, Egger M, Gluud LL *et al*. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;**336:**601–05.

[6] Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. *J Clin Epidemiol* 2011;**64:**1070–75.

[7] Savovic J, Jones HE, Altman DG *et al*. Influence of reported study design characteristics on intervention effect estimates from randomized controlled trials: combined analysis of meta-epidemiologic studies. *Ann Intern Med* 2012; In press.

[8] Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;**25:**3443–57.

[9] Rucker G, Carpenter JR, Schwarzer G. Detecting and adjusting for small-study effects in meta-analysis. *Biom J* 2011;**53:**351–68.

[10] Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;**21:** 1513–24.

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

[12] Welton NJ, Ades AE, Carlrin B, Altman DG, Sterne JA. Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Statist Soc A* 2009; **172:**119–36.

[13] Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *J R Statist Soc A* 2010;**173:** 613–29.

[14] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;**23:** 3105–24.

[15] Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;**17:**279–301.

[16] Salanti G, Dias S, Welton NJ, Ades AE, Golfinopoulos V, Kyrgiou M *et al*. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med* 2010;**29:** 2369–83.

[17] Franchini AJ, Dias S, Ades AE, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. *Res Syn Meth* 2012;**3:**142–160.

[18] Dias S, Welton NJ, Sutton AJ, Ades AE. A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. *NICE Decision Support Unit Evidence Synthesis Technical Support* Document 2, 2011.

[19] Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley & Sons, 2008.

[20] Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* 2009; **62:**857–64.

[21] Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med* 2009;**28:**1861–81.

[22] Moreno SG, Sutton AJ, Ades AE, Cooper NJ, Abrams KR. Adjusting for publication biases across similar interventions performed well when compared with gold standard data. *J Clin Epidemiol* 2011;**64:**1230–41.

[23] Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley & Sons, 2008.

[24] Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. *Crit Care* 2013;**17:**1.

[25] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. *J R Statist Soc B* 2002;**64:**583–639.

[26] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000;**10:**325–37.
[27] Abdullah AK, Khan S. Relative oral corticosteroid-sparing effect of 7 inhaled corticosteroids in chronic asthma: a meta-analysis. *Ann Allergy Asthma Immunol* 2008;**101:**74–81.
[28] Balk EM, Bonis PA, Moskowitz H *et al*. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;**287:**2973–82.
[29] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273:**408–12.
[30] Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med* 2007;**26:**2745–58.
[31] Hrobjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I *et al*. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012;**344:**e1119.
[32] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;**135:**982–89.
[33] Salanti G. Indirect and mixed-treatment comparison, network, or multiple- treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Meth* 2012;**3:**80.
[34] Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Res Synth Meth* 2012;**3:**161.
[35] Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. *Stat Med* 2013;**32:**51–66.

Published by Oxford University Press on behalf of the International Epidemiological Association
*International Journal of Epidemiology* 2013;**42**:1131–1133
doi:10.1093/ije/dyt137


# Commentary: Meta-epidemiology, meta-meta-epidemiology or network meta-epidemiology?

**Ludovic Trinquart,[1,2,3] Agnes Dechartres[1,4,5] and Philippe Ravaud[1,2,3,4,5]***

[1]INSERM U738, Paris, France, [2]Centre Cochrane Français, Paris, France, [3]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA, [4]Assistance Publique-Hôpitaux de Paris, Hôpital Hôtel-Dieu, Centre d'Épidémiologie Clinique, Paris, France and [5]Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Paris, France

*Corresponding author. Hôpital Hôtel-Dieu, Centre d'Épidémiologie Clinique. 1 Place du Parvis Notre-Dame, 75004 Paris, France. Email: philippe.ravaud@htd.aphp.fr

**Accepted** 14 June 2013

Identifying clinical trial characteristics influencing intervention effect estimates is crucial. Trials with such characteristics may lead to underestimating or overestimating true intervention effects. To assess risk of bias in randomized trials, the Cochrane Collaboration has developed a tool based on theoretical as well as empirical considerations regarding the impact of risk factors for bias.[1] Empirical evidence comes from meta-epidemiology. This approach involves use of a collection of meta-analyses to compare intervention effect estimates among trials with and without a particular characteristic. More recently, meta-meta-epidemiology, which combines data from several meta-epidemiological studies, has been developed.[2,3] In this issue of the journal, Chaimani *et al*.[4] propose network meta-epidemiology as an interesting new approach: meta-epidemiology in the framework of networks of trials, thus exploiting the assumption that the impact of risk factors is similar within networks.

Table 1 compares the methodological features of each approach. Each approach has pros and cons related to differences in data sources and assessment of risk factors. Meta-meta-epidemiology involves larger and probably more representative collections of meta-analyses than meta-epidemiology or network meta-epidemiology. In meta-epidemiology, an important restriction is that informative meta-analyses must include at least one trial with and one without the risk factor of interest. Moreover, a minimum number of trials per meta-analysis may be required, depending on how heterogeneity is modelled and whether multivariable analyses are undertaken. In network meta-epidemiology,