Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts and examples

Richard D. Riley*[1], Dan Jackson[2], Georgia Salanti[3,4], Danielle L. Burke[1], Malcolm Price[5], Jamie Kirkham[6], Ian R. White[2,7]

Contact details:

* corresponding author:

Professor of Biostatistics; e- mail: r.riley@keele.ac.uk;

Tel: +44 (0) 1782 733905 Fax: +44 (0) 1782 734719

[1] Research Institute for Primary Care and Health Sciences, Keele University, Staffordshire, UK. ST5 5BG

[2] MRC Biostatistics Unit, Cambridge, UK.

[3] Institute of Social and Preventive Medicine, University of Bern, Switzerland

[4] University of Ioannina School of Medicine, Ioanninga, Greece

[5] Institute of Applied Health Research, University of Birmingham, UK

[6] MRC North West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool, Liverpool, UK

[7] MRC Clinical Trials Unit at UCL, London, UK.

Word count: 5312

Acknowledgements:

Richard Riley, Dan Jackson, Malcom Price, Jamie Kirkham, and Ian White were supported by funding from a multivariate meta-analysis grant from the MRC Methodology Research Programme (grant reference number: MR/J013595/1). Daniel Jackson and Ian White were also supported by the Medical Research Council Unit (Programme number: U105260558). Georgia Salanti is supported by a Marie Skłodowska-Curie Fellowship (MSCA-IF-703254). Danielle Burke is funded by an NIHR School for Primary Care Research Post-Doctoral Fellowship. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We would like to thank the Editors and two reviewers for their constructive comments to improve the article upon revision.

Competing interests: We have read and understood the BMJ Group policy on declaration of interests and declare we have no competing interests

Contribution statement: RR conceived the article content and structure, with initial feedback from IW. RR applied the methods to the examples, with underlying data and software provided by IW, DJ and GS. RR wrote the first draft. All authors helped revise the paper, which included adding new text pertinent to their expertise, and refining the examples for the intended audience. RR revised the article, with further feedback from all other authors. RR is the guarantor.

Summary Points Box

- Meta-analysis methods combine quantitative evidence from related studies to produce results based on a whole body of research.
- Studies that do not provide *direct evidence* about a particular outcome or treatment comparison of interest are often discarded from a meta-analysis of that outcome or treatment comparison.
- Multivariate and network meta-analysis methods simultaneously analyse multiple outcomes and multiple treatments, respectively. This allows more studies to contribute toward each outcome and treatment comparison.
- Summary results for each outcome now depend on correlated results from other outcomes, and summary results for each treatment comparison now incorporate indirect evidence from related treatment comparisons, in addition to any direct evidence.
- This often leads to a gain in information which can be quantified by the Borrowing of Strength statistic, *BoS* (the percentage reduction in the variance of a summary result that is due to correlated or indirect evidence).
- Under a missing at random assumption, a multivariate meta-analysis of multiple outcomes is most beneficial when the outcomes are highly correlated and the percentage of studies with missing outcomes is large.
- Network meta-analyses gain information through a consistency assumption, which should be evaluated in each network where possible. There is usually low power to detect inconsistency, which arises when effect modifiers are systematically different in the subsets of trials providing direct and indirect evidence.
- Network meta-analysis allows multiple treatments to be compared and ranked based on their summary results. However, focusing on the probability of being ranked first is potentially misleading: a treatment ranked first may also have a high probability of being ranked last, and its benefit over other treatments may be of little clinical value.
- Novel network meta-analysis methods are emerging to use individual participant data, to evaluate dose, to incorporate 'real-world' evidence from observational studies, and to relax the consistency assumption by allowing summary inferences whilst accounting for inconsistency effects.

Standfirst

Organisations such as the National Institute for Health and Care Excellence (NICE) require evidence synthesis of existing studies to inform their decisions, for example about the best available treatments with respect to multiple efficacy and safety outcomes. However, relevant studies may not provide direct evidence about all the treatments or outcomes of interest. Multivariate and network meta-analysis methods provide a framework to address this, using correlated and/or indirect evidence from such studies alongside any direct evidence. In this article, Riley and colleagues describe the key concepts and assumptions of these methods, outline how correlated and indirect evidence arises, and illustrate the contribution of such evidence in real clinical examples involving multiple outcomes and multiple treatments.

Introduction and rationale

Meta-analysis methods combine quantitative evidence from related studies to produce results based on a whole body of research. As such, meta-analyses are an integral part of evidence based medicine and clinical decision-making, for example to guide which treatment should be recommended for a particular condition. The majority of meta-analyses are based on combining results (e.g. treatment effect estimates) extracted from study publications or obtained directly from study authors. Unfortunately, relevant studies may not evaluate the same sets of treatments and outcomes, which creates problems for meta-analysis. For example, in a meta-analysis of 28 trials to compare eight thrombolytic treatments after acute myocardial infarction, it is unrealistic to expect every trial to compare all eight treatments;[1] in fact a different set of treatments was examined in each trial, with the maximum number of trials per treatment was only eight.[1] Similarly, relevant clinical outcomes may not always be available. For example, in a meta-analysis to summarise the prognostic effect of progesterone receptor status in endometrial cancer, four studies provided results for both cancer-specific survival (CSS) and progression-free survival (PFS), but other studies provided results for only CSS (2 studies) or only PFS (11 studies).[2]

Studies that do not provide *direct evidence* about a particular outcome or treatment of interest are often excluded from a meta-analysis evaluating that outcome or treatment. This is unwelcome, especially if their participants are otherwise representative of the population, clinical settings and condition of interest. Research studies require considerable costs and time, and involve precious patient involvement, and simply discarding them could be viewed as research waste.[3-5] Statistical models for *multivariate* and *network meta-analysis* address this by simultaneously analysing multiple outcomes and multiple treatments, respectively. This allows more studies to contribute toward each outcome and treatment comparison. Furthermore, in addition to using direct evidence,

the summary result for each outcome now depends on correlated results from related outcomes, and the summary result for each treatment comparison now incorporates indirect evidence from related treatment comparisons.[6 7] The rationale is that by observing the related evidence we learn something about the missing direct evidence of interest, and thus gain some information that is otherwise lost; a concept sometimes known statistically as 'borrowing strength'.[6 8]

Multivariate and, in particular, network meta-analyses are increasingly prevalent in clinical journals. For example, a review up to April 2015 identified 456 network meta-analyses of randomised trials evaluating at least four different interventions.[9] Only six of these 456 were published before 2005, and 103 were published in 2014 alone, emphasising a dramatic increase in uptake in the last ten years (Figure 1(a)). The BMJ has published more than any other journal (28; 6.1%). Methodology and tutorial articles about network meta-analysis have also risen in number, from less than five in 2005 to over 30 per year since 2012 (Figure 1(b)).[10]

Here we explain the key concepts, methods, and assumptions of multivariate and network meta-analysis, building on previous pieces in The BMJ.[11 12 13] We begin by describing the use of correlated effects within a multivariate meta-analysis of multiple outcomes, and then consider the use of indirect evidence within a network meta-analysis of multiple treatments. We also highlight two statistics ('*BoS*' and '*E*') that summarise the extra information gained, and consider key assumptions, challenges and novel extensions. Real examples are embedded throughout.

Correlated effects and multivariate meta-analysis of multiple outcomes

> *"Many clinical studies have more than one outcome variable; this is the norm rather than the exception. These variables are seldom independent and so each must carry some information about the others. If we can use this information, we should."*
>
> Martin Bland [14]

Many clinical outcomes are correlated with each other, such as a hypertensive patient's systolic and diastolic blood pressure, a migraine sufferer's level of pain and nausea, and a cancer patient's disease-free and overall survival times. Such correlation at the individual level will lead to correlation between effects at the population (study) level. For example, in a randomised trial of anti-hypertensive treatment, the estimated treatment effects for systolic and diastolic blood pressure are likely to be highly correlated. Similarly, in a cancer cohort study the estimated prognostic effects of a biomarker are likely to be highly correlated for disease-free survival and overall survival. Correlated effects also arise in many other situations, for example when there are multiple time-points

(longitudinal data);[15] multiple biomarkers and genetic factors that are interrelated;[16] multiple effect sizes corresponding to overlapping sets of adjustment factors;[17] multiple measures of accuracy or performance (e.g. in regard to a diagnostic test or prediction model),[18] and multiple measures of the same construct (e.g. scores from different pain scoring scales, or biomarker values from different laboratory measurement techniques[19]). We broadly refer to these as multiple correlated *outcomes* in this article.

As Bland notes,[14] correlation amongst outcomes is potentially informative and worth using. A multivariate meta-analysis addresses this by analysing all correlated outcomes *jointly*. This is usually achieved by assuming multivariate normal distributions,[7 20] and generalises standard ('univariate') meta-analysis methods described previously in The BMJ.[12] Note that the outcomes are not amalgamated into a single outcome; the multivariate approach still produces a *distinct* summary result for each outcome. However, the correlation amongst the outcomes is now incorporated, and this brings two major advantages compared to a univariate meta-analysis of each outcome separately. Firstly, the incorporation of correlation enables each outcome's summary result to make use of the data for *all* outcomes. Secondly, we can now include studies that do not report all the outcomes of interest.[21] This allows more studies and evidence to be included, and consequently can lead to more precise conclusions (narrower confidence intervals). More technical details and software options are provided in Supplementary Material 1. [22 23 24 25] We now illustrate the key concepts through two examples.

Example 1: Prognostic effect of progesterone for cancer specific survival in endometrial cancer

In the aforementioned endometrial cancer example, prognostic results for CSS are missing in 11 studies (1412 patients) that provide results for PFS. A traditional univariate meta-analysis for CSS simply discards these 11 studies, but they are retained in a multivariate analysis of PFS and CSS, which uses their strong positive correlation (about +0.8). This leads to important differences in summary results, as shown for CSS in the forest plot of Figure 2. The univariate meta-analysis for CSS includes just the six studies with direct evidence and gives a summary hazard ratio (HR) of 0.61 (95%: 0.38 to 1.00; I^2 = 70%), with the confidence interval just crossing the value of no effect. The multivariate meta-analysis includes 17 studies and gives a summary HR for CSS of 0.48 (95% CI: 0.29 to 0.79), with a narrower confidence interval and stronger evidence that progesterone is prognostic for CSS. The latter result is also more similar to the prognostic effect for PFS (summary HR = 0.43, 95% CI: 0.26 to 0.71, from multivariate meta-analysis), as perhaps one might expect.

Example 2: Plasma fibrinogen concentration as a risk factor for CVD

The Fibrinogen Studies Collaboration examine whether plasma fibrinogen concentration is an independent risk factor for cardiovascular disease (CVD) using data from 31 studies.[17] All 31 studies allowed a *partially* adjusted hazard ratio to be obtained, where the hazard ratio for fibrinogen was adjusted for the same core set of known risk factors, including age, smoking, BMI and blood pressure. However, a more '*fully*' adjusted hazard ratio, additionally adjusted for cholesterol, alcohol consumption, triglycerides and diabetes, was only calculable in 14 studies. When the partially and 'fully' adjusted estimates are plotted in these 14 studies, there is a strong positive correlation (almost +1, i.e. a near perfect linear association) between them (Figure 3).

A standard (univariate) random effects meta-analysis of just the direct evidence from 14 trials gives a summary 'fully' adjusted HR of 1.31 (95% CI: 1.22 to 1.42; I^2 = 29%), which indicates that a 1 g/L increase in fibrinogen levels is associated, on average, with a 31% relative increase in the hazard of CVD. However, a multivariate meta-analysis of partially and 'fully' adjusted results incorporates information from all 31 studies, and thus an additional 17 studies (>70000 patients), to utilise their large correlation (close to +1). This produces the same 'fully' adjusted summary HR of 1.31, but gives a more precise confidence interval (1.25 to 1.38) due to the extra information gained. A forest plot is shown in supplementary material 2.

Indirect evidence and network meta-analysis of multiple treatments

Let us now consider the evaluation of multiple treatments. A meta-analysis that evaluates a particular treatment comparison (e.g. treatment A versus B) using only direct evidence is known as a 'pair-wise meta-analysis'. When the set of treatments differs across trials, this approach may greatly reduce the number of trials per meta-analysis, and makes it hard to formally compare more than two treatments. A network meta-analysis addresses this by synthesising all trials in the same analysis whilst utilising *indirect evidence*.[22 26 27]

Consider a simple network meta-analysis of three treatments (A, B and C) evaluated in previous randomised trials. Assume that the relative treatment effect (i.e. the treatment contrast) of A versus B is of key interest, and that some trials compare A and B directly. However, there are also other trials of A versus C and other trials of B versus C, which provide no direct evidence of the benefit of A versus B, as they did not examine both A and B. Indirect evidence of A versus B can still be obtained from these trials under the so-called "consistency" assumption that, on average across all trials regardless of the treatments compared, the

Treatment contrast of A versus B

= (treatment contrast of A versus C) – (treatment contrast of B versus C),

where 'treatment contrast' is, for example, a log relative risk, log odds ratio, log hazard ratio or mean difference. This relationship will always hold *exactly* within any randomised trial where A, B and C are all examined. However, it is plausible that it will also hold (on average) across those trials that only compare a reduced set of treatments, if their clinical and methodological characteristics (such as quality, length of follow-up, case-mix) are similar in each subset (here, A versus B, A versus C, and B versus C trials). In this situation, the benefit of A versus B can be inferred from the indirect evidence from comparing trials of just A versus C with trials of just B versus C, in addition to the direct evidence coming from trials of A versus B (Figure 4).

There are different options for specifying a network meta-analysis model under this consistency assumption, depending on the type of data available. If there are only two treatments (i.e. one treatment comparison) per trial, then the simplest approach is a standard meta-regression, which models the treatment effect estimates across trials in relation to a reference treatment. The choice of reference treatment is arbitrary, and makes no difference to the meta-analysis results. This can be extended to a multivariate meta-regression to accommodate trials with 3 or more groups (often called 'multi-arm trials').[28 29] Rather than modelling treatment effect estimates directly, for a binary outcome it is more common to use a logistic regression framework to model the numbers and events available for each treatment group (arm) directly. Similarly, a linear regression or Poisson regression could be used to directly model continuous outcomes and rates in each group in each trial. When doing so, it is important to maintain the randomisation and clustering of patients within trials,[28] and to incorporate random effects to allow for between-trial heterogeneity in the magnitude of treatment effects.[12] Supplementary Material 1 gives more technical details (and software options[27 30]) for network meta-analysis, and fuller statistical explanation is given elsewhere.[28]

After estimation of a network meta-analysis, a summary result is obtained for each treatment relative to the chosen reference treatment. Subsequently, other comparisons (treatment contrasts) are then derived using the consistency relationship. For example, if C is the reference treatment in a network meta-analysis of a binary outcome, then the summary log odds ratio (logOR) for A versus B is obtained by the difference in the summary logOR estimate for A versus C and the summary logOR estimate for B versus C. We now illustrate the key concepts through an example.

Example 3: Comparison of eight thrombolytic treatments after acute myocardial infarction

In the aforementioned thrombolytics meta-analysis,[1] the aim was to estimate the relative efficacy of eight competing treatments in reducing the odds of mortality by 30-35 days; these treatments are labelled as A to H for brevity (for full names see below Figure 5). A version of this dataset containing seven treatments was previously introduced in the BMJ by Caldwell et al.,[13] and our investigations below extend this work.

With eight treatments, 28 pair-wise comparisons of potential interest; however, only 13 of these were directly reported in at least one trial. This is shown by the network of trials (Figure 5), where each node is a particular treatment, and a line connects two nodes when at least one trial directly compares the two respective treatments. For example, a direct comparison of C versus A is available in eight trials, whilst a direct comparison of F versus A is only available in one trial. With such discrepancy in the amount of direct evidence available for each treatment, and between each pair of treatments, it is hugely problematic to compare the eight treatments using only standard (univariate) pair-wise meta-analysis methods.

Therefore, using the number of patients and deaths by 30-35 days in each treatment group, we applied a network meta-analysis via a multivariate random effects meta-regression model, to obtain the summary odds ratios for treatments B to H versus A, and subsequently all other contrasts.[27 29] This allowed all 28 trials to be incorporated and all eight treatments to be compared simultaneously, utilising direct evidence and also indirect evidence propagated through the network via the consistency assumption. The choice of reference group does not change the results, which are displayed in Figure 6 and supplementary material 3. The indirect evidence has an important impact on some treatment comparisons. For example, the summary treatment effect of H versus B in the network meta-analysis of all 28 trials (OR 1.19, 95% CI: 1.06 to 1.35) is substantially different from a standard pair-wise meta-analysis of two trials (summary OR 3.87; 95% CI: 1.74 to 8.58).

Ranking treatments

Following a network meta-analysis it is helpful to rank treatments according to their effectiveness. This process usually, though not always,[31] requires using simulation or resampling methods.[27 29 32] These use thousands of samples from the (approximate) distribution of summary treatment effects, to identify the percentage of samples (probability) that each treatment has the most (or least) beneficial effect. Figure 7(a) shows the probability that each thrombolytic treatment was ranked most effective out of all treatments, and similarly second, third, and so on down to the least

effective. Treatment G has the highest probability (51.7%) of being the most effective at reducing the odds of mortality by 30-35 days, followed by treatment E (21.5%) and B (18.3%).

Focusing on the probability of being ranked first is potentially misleading: a treatment ranked first may also have a high probability of being ranked last,[33] and its benefit over other treatments may be of little clinical value. In our example, treatment G has the highest probability of being most effective, but the summary effect for G is very similar to that for B and E, and their difference is unlikely to be clinically important. Furthermore, treatment G is also fourth most likely to be the least effective (14.4%), reflecting a large summary effect with a wide confidence interval. In contrast treatments B, E and F have very low probability (close to 0%) of being least effective. Thus, a treatment may have the highest probability of being ranked first, when actually there is no strong evidence (beyond chance) that it is better than other available treatments. To illustrate this further, let us add to the thrombolytics network a hypothetical new drug, called *Brexitocin*, for which no direct or indirect evidence exists. Given the lack of evidence, *Brexitocin* essentially has a 50% chance of being the most effective treatment but also a 50% chance of being the least effective.

To help address this, the mean rank and the Surface Under the Cumulative RAnking curve (SUCRA) are useful.[34 35] The mean rank gives the average ranking place for each treatment. The SUCRA is the area under a line plot of the cumulative probability over ranks (from most effective to least effective) (Figure 7(b)), and is just the mean rank scaled to be between 0 and 1. A similar measure is the P-score.[31]F or the thrombolytic network (now excluding *Brexitocin*), Treatments B and E have the best mean ranks (2.3 and 2.6, respectively), followed by treatment G (3.0). Thus, although treatment G had the highest probability of being ranked first, based on the mean rank it is now in third place.

Quantifying the information gained from correlated or indirect evidence

Copas et al. (submitted) propose that, in comparison to a multivariate or network meta-analysis with the same magnitude of between-trial heterogeneity, a standard (univariate) meta-analysis of just the direct evidence is similar to throwing away $100 \times (1 - E)$% of the available studies. The efficiency (E) is defined by,

$$E = \frac{\text{variance of summary result based on direct and related evidence}}{\text{variance of summary result based on only direct evidence}}$$

where 'related evidence' refers to either indirect or correlated evidence (or both), and the variance relates to the original scale of the meta-analysis (so typically the log relative risk, log odds ratio, log hazard ratio, or mean difference). For example, if *E* = 0.9 then a standard meta-analysis is similar to throwing away 10% of available studies and patients (and events).

Let us also define *n* as the number of available studies with direct evidence (i.e. those that would contribute toward a standard meta-analysis). Then, the extra information gained toward a particular summary meta-analysis result by using indirect or correlated evidence can also be considered similar to having found direct evidence from a further $n \times \frac{(1-E)}{E}$ studies of a similar size to the *n* trials. For example, if there are nine studies providing direct evidence about an outcome for a standard univariate meta-analysis and *E* = 0.9, then the advantage of using a multivariate meta-analysis is like finding direct evidence for that outcome from a further $9 \times \frac{(1-0.9)}{0.9} = 1$ study. We thus gain the considerable time, effort and money invested in about one research study.

Jackson et al. also propose the 'borrowing of strength' (*BoS*) statistic,[8] which can be calculated for each summary result within a multivariate or network meta-analysis by

$$BoS = 100 \times (1 - E)\%.$$

BoS provides the percentage reduction in the variance of a summary result that is due to (borrowed from) correlated or indirect evidence. An equivalent way of interpreting *BoS* is the percentage weight in the meta-analysis that is given to the correlated or indirect evidence.[8] For example, in a network meta-analysis, a *BoS* of 0% indicates that the summary result is based only on direct evidence, whereas a *BoS* of 100% indicates that it is based entirely on indirect evidence. Riley et al. show how to derive percentage study weights for multi-parameter meta-analysis models, including network and multivariate meta-analysis.{Riley, 2017 #1486}

Application to the examples

Let us revisit our three examples. In the fibrinogen example, the summary 'fully' adjusted HR has a large *BoS* of 53%, indicating that the correlated evidence (from the partially adjusted results) contributes 53% of the total weight toward the summary result. The efficiency (*E*) is 0.47, and thus using the correlated evidence is equivalent to having found 'fully' adjusted results from an additional $14 \times \frac{(1-0.47)}{0.47} \approx 16$ studies.

In the progesterone example, *BoS* is 33% for CSS indicating that using the PFS results reduces the variance of the summary log hazard ratio for CSS by 33%. This corresponds to an *E* of 0.67, and the information gained from the multivariate meta-analysis can be considered similar to having found CSS results from an additional $6 \times \frac{(1-0.67)}{0.67} \approx 3$ studies.

For the thrombolytics meta-analysis, *BoS* is shown in Figure 6 for each treatment comparison where there was direct evidence for at least one trial. It is often large. For example, the comparison of H versus B has a *BoS* of 97.8%, as there are only two trials with direct evidence. This is similar to having found direct evidence for H versus B from an additional $2 \times \frac{(1-0.022)}{0.022} \approx 89$ trials of similar size to those existing two trials. *BoS* is 0% for E versus B, as there was no indirect evidence toward this comparison (Figure 6). For comparisons not shown in Figure 6, such as C versus B, *BoS* was 100% because there was no direct evidence. The percentage weight (contribution) of each study is shown in Supplementary material 3.

Challenges and assumptions of multivariate or network meta-analysis

Our three examples demonstrate the potential value of multivariate and network meta-analysis, and other benefits are discussed elsewhere.[15 20 36] However, the approaches do have issues.

The benefits of a multivariate meta-analysis may be small

> *"… multivariate and univariate models generally give similar point estimates, although the multivariate models tend to give more precise estimates. It is unclear, however, how often this added precision will qualitatively change conclusions of systematic reviews".*
>
> Trikalinos et al.[37]

This argument, based on empirical evidence,[37] might be levelled at the fibrinogen example. Although there was considerable gain in precision from using multivariate meta-analysis (*BoS* = 53%), fibrinogen was clearly identified as a risk factor for CVD in both univariate and multivariate analyses, and thus conclusions did not change. A counter-view is that this is in itself useful to know.

The potential importance of a multivariate meta-analysis of multiple outcomes is greatest when *BoS* and *E* are large, which is more likely when:

- the proportion of studies *without* direct evidence for an outcome of interest is large;
- results for other outcomes are available in studies where an outcome of interest is not reported; and

- the magnitude of correlation amongst outcomes is large (e.g. > 0.5 or < -0.5), either within-studies or between-studies.

In our experience, *BoS* and *E* are usually greatest in a network meta-analysis of multiple treatments; that is, more information is usually gained about multiple treatments via the consistency assumption than is gained about multiple outcomes via correlation. A multivariate meta-analysis of multiple outcomes is best reserved for a set of *highly* correlated outcomes, as otherwise *BoS* and *E* are usually small. Such outcomes should be identified and specified in advance of analysis, for example using clinical judgement and statistical knowledge, so as to avoid data dredging across different sets of outcomes. A multivariate meta-analysis of multiple outcomes is also best reserved for a situation with *missing* outcomes (at the study-level), as anecdotal evidence suggests that *BoS* for an outcome is approximately bounded by the percentage of missing data for that outcome. For example, in the fibrinogen example the percentage of trials with a missing fully adjusted outcome is 55% (= 100% × 17/31), and thus the multivariate approach is flagged as worthwhile as *BoS* could be as high as 55% for the fully adjusted pooled result. As discussed, the actual *BoS* was 53% and thus very close to 55%, due to the near perfect correlation between partially and fully adjusted effects. In contrast, in situations with complete data or a low percentage of missing outcomes, *BoS* (and thus a multivariate meta-analysis) is unlikely to be important. Also, multivariate meta-analysis cannot handle trials that do not report *any* of the outcomes of interest. Therefore, although it can reduce the impact of selective outcome reporting in published trials, it cannot reduce the impact of non-publication of entire trials (publication bias).

If a formal comparison of correlated outcomes is of interest (e.g. to estimate the difference between the treatment effects on systolic and diastolic blood pressure), then this should always be done in a multivariate framework regardless of the amount of missing data, in order to account for correlations between outcomes and thus avoid erroneous confidence intervals and p-values.[38] Similarly, a network meta-analysis of multiple treatments is preferable even if all trials examine all treatments, as we require a single analysis framework for estimating and comparing the effects of each treatment.

Model specification and estimation is non-trivial

Even when *BoS* is anticipated to be large, challenges may remain.[20] Multivariate and network meta-analysis models are often complex, and achieving convergence (i.e. reliable parameter estimates) may require simplification (e.g. common between-study variance terms for each treatment contrast;

multivariate normality assumption), which may be open to debate.[20 39 40] For example, in a multivariate meta-analysis of multiple outcomes, convergence and estimation problems increase as the number of outcomes (and hence unknown parameters) increase, and so applications beyond two or three outcomes are rare. Specifically, unless IPD are available,[41] there can be problems obtaining and estimating correlations amongst outcomes;[42 43] possible solutions include a Bayesian framework utilising prior distributions for unknown parameters to bring in external information. [44-46] [46-48]

Benefits arise under assumptions

> *"But borrowing strength builds weakness. It builds weakness in the borrower because it reinforces dependence on external factors to get things done."*
>
> Stephen Covey [49]

This quote relates to qualities needed for an effective leader, but is pertinent here too. The benefits of multivariate and network meta-analysis depend on missing study results being ***missing at random***.[50] We are assuming that the relationships that we *do* observe in some trials are transferable to other trials where they are unobserved. For example, in a multivariate meta-analysis of multiple outcomes the observed linear association (correlation) of effects for pairs of outcomes (both within-studies and between-studies) is assumed to be transferable to other studies where only one of the outcomes is available. This relationship is also used to justify surrogate outcomes,[51] but often receives criticism and debate therein.[52] Missing *not* at random may be more appropriate when results are missing due to selective outcome reporting,[53] or selective choice of analyses.[54] A multivariate approach may still reduce selective reporting biases in this situation,[36] but not completely.

In a network meta-analysis of multiple treatment comparisons, the missingness assumption is also known as *transitivity*;[55 56] it implies that the *relative* effects of three or more treatments observed directly in some trials would be the same in other trials where they are unobserved. Based on this, the consistency assumption then holds. When the direct and indirect evidence disagree, this is known as inconsistency (incoherence). A recent review by Veroniki et al. found that about one in eight network meta-analyses show inconsistency as a whole,[57] similar to an earlier review.[58]

How do we examine inconsistency between direct and indirect evidence?

Treatment effect modifiers relate to methodological or clinical characteristics of the trials that influence the magnitude of treatment effects, and may include follow-up length, outcome

definitions, study quality (risk of bias), analysis and reporting standards (including risk of selective reporting), and the patient-level characteristics.[58-61] When such effect modifiers are systematically different in the subsets of trials providing direct and indirect evidence, this causes genuine inconsistency. Thus, before undertaking a network meta-analysis it is important to select only those trials relevant for the population of clinical interest, and to then identify any systematic differences in those trials providing different comparisons. For example, in the thrombolytics network, are trials of A versus C and A versus H systematically different from trials of C versus H in terms of potential effect modifiers?[62] If so, inconsistency is likely and so a network meta-analysis approach is best avoided.

It may be difficult to gauge the potential for inconsistency in advance of a meta-analysis. Therefore, following any network meta-analysis, inconsistency should be examined statistically, though unfortunately this is often not done.[63] The consistency assumption can be examined for each treatment comparison where there is direct and indirect evidence (seen as a closed loop within the network plot):[57 64 65] here the 'separating indirect from direct evidence' [65] approach (sometimes called 'node-splitting' or 'side-splitting') involves estimating the direct and indirect evidence, and comparing the two. The consistency assumption can also be examined across the whole network using 'design-by-treatment interaction' models,[29 66] which allow an overall significance test for inconsistency. If evidence of inconsistency is found, explanations should be sought: for example, whether inconsistency arises from particular studies with a different design or at a higher risk of bias.[55] The network models could then be extended to include suitable explanatory covariates or reduced to exclude certain studies.[62] If inconsistency remains unexplained, then the inconsistency terms may instead be modelled as random effects with mean zero, thus allowing overall summary estimates allowing for unexplained inconsistency.[67-69] Other approaches for modelling inconsistency have been proposed,[64] and we anticipate further developments in this area over the coming years. However, often there is low power to detect genuine inconsistency.[70]

In the thrombolytics example, the 'separating indirect from direct evidence' approach found no significant inconsistency except for H versus B, visible in Figure 6 as the discrepancy between "Study 22", "Study 23" and "All studies" under the subheading "H vs B". However, when we applied the 'design-by-treatment interaction' model there was no evidence of overall inconsistency. If the H versus B studies differed in design from the other studies then it might be reasonable to exclude them from the network, but otherwise an overall inconsistency model (with inconsistency terms included as random effects) may provide the best treatment comparisons.

Novel extensions and hot topics

Incorporation of both multiple treatments and multiple outcomes

Previous examples considered either multiple outcomes or multiple treatments. However, there is growing interest in accommodating both together, in order to help identify the best treatment across multiple clinically relevant outcomes.[71-76] This is achievable, but challenging due to the extra complexity of the statistical models required. For example, Efthimiou et al.[72] perform a network meta-analysis of 68 studies comparing 13 active antimanic drugs and placebo for acute mania. Two primary outcomes of interest were *efficacy* (defined as the proportion of patients with at least a 50% reduction in manic symptoms from baseline to week 3) and *acceptability* (defined as the proportion of patients with treatment discontinuation before 3 weeks). These are likely to be negatively correlated (as patients often discontinue treatment due to lack of efficacy), so the authors extend a network meta-analysis framework to jointly analyse these outcomes and account for their correlation (estimated to be about -0.5). This is especially important as 19 of the 68 studies provided data on only one of the two outcomes. Compared to considering each outcome separately, this approach produces narrower confidence intervals for summary treatment effects and has an impact on the relative ranking of some of the treatments (Supplementary Material 4). In particular, Carbamazepine ranks as the most effective treatment in terms of response when considering outcomes separately, but falls to fourth place when accounting for their correlation.

Accounting for dose and class

Standard network meta-analysis makes no allowance for similarities between treatments. When some treatments represent different doses of the same drug, network meta-analysis models may be extended to incorporate sensible dose-response relationships.[77] Similarly, when the treatments can be grouped into multiple classes, network meta-analysis models may be extended to allow treatments in the same class to have more similar effects than treatments in different classes.[78]

Use of individual participant data (IPD)

Network meta-analysis using aggregate (published) data is convenient, but sometimes published reports are inadequate for this purpose: for example, if outcome measures are differently defined, or if interest lies in treatment effects within subgroups. In these cases it may be valuable to collect IPD.[79] As such, methods for network meta-analysis of IPD are emerging.[60 80-85] A major advantage is that these allow the inclusion of participant-level covariates, which is important if these are effect modifiers that would otherwise cause inconsistency in the network.

Inclusion of 'real-world' evidence

There is growing interest in using 'real-world' evidence from non-randomized studies in order to corroborate findings from randomised trials, and to increase the evidence being used toward decision-making. Network meta-analysis methods are thus being extended for this purpose,[86] and a recent overview is given by Efthimiou et al.,[87] who emphasise the importance of ensuring compatibility of the different pieces of evidence, for each treatment comparison.

Cumulative network meta-analysis

Créquit et al.[88] show that the amount of randomized evidence covered by existing systematic reviews of competing second-line treatments for advanced non-small cell lung cancer was always substantially incomplete, with 40 % or more of treatments, treatment comparisons, and trials missing. To address this, they recommend a new paradigm "by switching: 1) from a series of standard meta-analyses focused on specific treatments (many treatments being not considered) to a single network meta-analysis covering all treatments; and 2) from meta-analyses performed at a given time and frequently out-of-date to a cumulative network meta-analysis systematically updated as soon as the results of a new trial become available." They latter is referred to as a "live cumulative network meta-analysis", and the various steps, advantages and challenges of this approach warrant further consideration.[88] A similar concept is the Framework for Adaptive MEta-analysis (FAME), which requires knowledge of ongoing trials and suggests timing meta-analysis updates to coincide with new publications.[89]

Quality assessment and reporting

Finally, we encourage quality assessment of network meta-analysis according to the guidelines of Salanti et al,[90] and clear reporting of results using the PRISMA-NMA guidelines.[91] The latter may be enhanced by the presentation of percentage study weights according to recent proposals,[8 92] to reveal the contribution of each study toward the summary treatment effects.

Conclusions

Statistical methods for multivariate and network meta-analysis use correlated and indirect evidence alongside direct evidence, and here we have highlighted their advantages and challenges. Table 1 summarises the rationale, benefits and potential pitfalls of the two approaches. Core outcome sets and data sharing will hopefully reduce the issue of missing direct evidence,[61 79 93] but are unlikely to

resolve it completely. Thus, to combine indirect and direct evidence in a coherent framework, we expect applications of, and methodology for, multivariate and network meta-analysis to continue to grow in the coming years.[9 94]

Reference List

1. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006;101(474):447-59.
2. Zhang Y, Zhao D, Gong C, Zhang F, He J, Zhang W, Zhao Y, Sun J. Prognostic role of hormone receptors in endometrial cancer: a systematic review and meta-analysis. *World journal of surgical oncology* 2015;13:208.
3. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, Al-Shahi Salman R, Chan AW, Glasziou P. Biomedical research: increasing value, reducing waste. *Lancet* 2014;383(9912):101-4.
4. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383(9913):267-76.
5. Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, Krumholz HM, Ghersi D, van der Worp HB. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383(9913):257-66.
6. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15(24):2733-49.
7. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21(4):589-624.
8. Jackson D, White IR, Price M, Copas J, Riley RD. Borrowing of strength and study weights in multivariate and network meta-analysis. *Stat Methods Med Res* 2017 (in-press).
9. Petropoulou M, Nikolakopoulou A, Veroniki AA, Rios P, Vafaei A, Zarin W, Giannatsi W, Sullivan S, Tricco A, Chaimani A, Egger M, Salanti G. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *J Clin Epidemiol* 2017; 82: 20-28.
10. Efthimiou O, Debray TP, van Valkenhoef G, Trelle S, Panayidou K, Moons KG, Reitsma JB, Shang A, Salanti G, GetReal Methods Review G. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods* 2016;7(3):236-63.
11. Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ* 2013;346:f2914.
12. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342:d549.
13. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331(7521):897-900.
14. Bland JM. Comments on 'Multivariate meta-analysis: Potential and promise' by Jackson et al. *Statistics in Medicine* 2011;30:2502-03.
15. Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials* 2009;6(1):16-27.
16. Thompson JR, Minelli C, Abrams KR, Tobin MD, Riley RD. Meta-analysis of genetic studies using Mendelian randomization--a multivariate approach. *Stat Med* 2005;24(14):2241-54.
17. Collaboration FS. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Stat Med* 2009;28(8):1218-37.
18. Snell KI, Hua H, Debray TP, Ensor J, Look MP, Moons KG, Riley RD. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol* 2016;69:40-50.
19. Riley RD, Elia EG, Malin G, Hemming K, Price MP. Multivariate meta-analysis of prognostic factor studies with multiple cut-points and/or methods of measurement. *Stat Med* 2015;34(17):2481-96.
20. Jackson D, Riley RD, White IR. Multivariate meta-analysis: potential and promise. *Stat Med* 2011;30:2481-98.

21. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ* 2013;346:e8668.
22. White IR. Multivariate random-effects meta-regression: Updates to mvmeta. *The Stata Journal* 2011;11:255-70.
23. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med* 2012;31:3821-39.
24. Jackson D, White IR, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Stat Med* 2012;31(29):3805-20.
25. Jackson D, White IR, Riley RD. A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *Biom J* 2013;55(2):231-45.
26. 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. *Research Synthesis Methods* 2012;3(2):80-97.
27. White IR. Network meta-analysis. *The Stata Journal* 2015;15:951-85.
28. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;17(3):279-301.
29. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Method* 2012;3: :111-25.
30. Rucker G, Schwarzer G. netmeta: An R package for network meta analysis. *The R Project website* 2013.
31. Rucker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015;15:58.
32. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23(20):3105-24.
33. Trinquart L, Attiche N, Bafeta A, Porcher R, Ravaud P. Uncertainty in Treatment Rankings: Reanalysis of Network Meta-analyses of Randomized Trials. *Ann Intern Med* 2016;164(10):666-73.
34. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One* 2013;8(10):e76654.
35. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011;64(2):163-71.
36. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Stat Med* 2012;31(20):2179-95.
37. Trikalinos TA, Hoaglin DC, Schmid CH. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. *Stat Med* 2014;33(9):1441-59.
38. 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* 2007;26(1):78-97.
39. Nikoloulopoulos AK. A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution. *Stat Med* 2015; 34(29):3842-65.
40. Senn S, Gavini F, Magrez D, Scheen A. Issues in performing a network meta-analysis. *Stat Methods Med Res* 2013;22(2):169-89.
41. Riley RD, Price MJ, Jackson D, Wardle M, Gueyffier F, Wang J, Staessen JA, White IR. Multivariate meta-analysis using individual participant data. *Res Synth Method* 2015;6 157-74.
42. 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(1):3.
43. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *JRSS Series A* 2009;172(4):789-811.

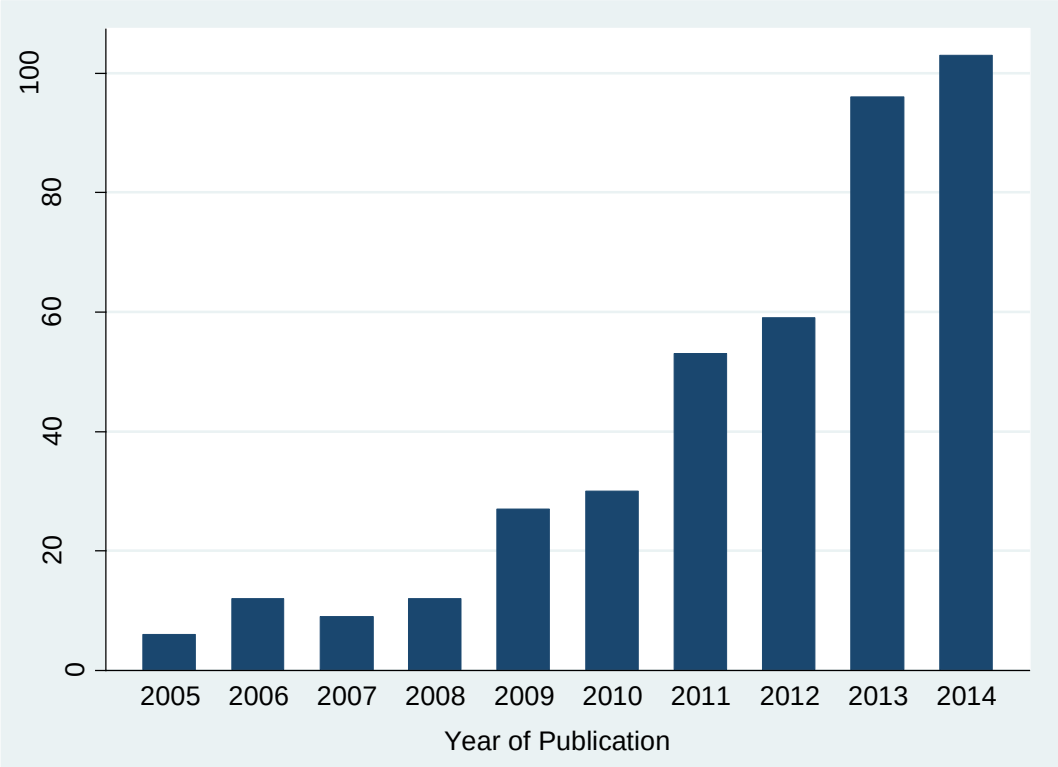
44. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2008;9(1):172-86.
45. Wei Y, Higgins JP. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Stat Med* 2013;32(7):1191-205.
46. Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DP, Abrams KR. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. *Stat Med* 2013;32(22):3926-43.
47. Wei Y, Higgins JP. Bayesian multivariate meta-analysis with multiple outcomes. *Stat Med* 2013;32(17):2911-34.
48. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int J Epidemiol* 2012;41(3):818-27.
49. Covey SR. The 7 Habits of Highly Effective People Personal Workbook: Simon & Schuster UK 2008.
50. Seaman S, Galati J, Jackson D, Carlin J. What Is Meant by "Missing at Random"? *Statistical Science* 2013;28(2):257-68.
51. Nixon RM, Duffy SW, Fender GR. Imputation of a true endpoint from a surrogate: application to a cluster randomized controlled trial with partial information on the true endpoint. *BMC Medical Research Methodology* 2003;3(1):17.
52. D'Agostino RB. Debate: The slippery slope of surrogate outcomes. *Trials* 2000;1(2):76.
53. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.
54. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, Williamson PR, Kirkham JJ. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* 2014;11(6):e1001666.
55. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013;159(2):130-7.
56. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health* 2014;17(2):157-73.
57. Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. *Int J Epidemiol* 2013;42(1):332-45.
58. Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, Holland R, Chen YF, Glenny AM, Deeks JJ, Altman DG. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ* 2011;343:d4909.
59. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC medicine* 2013;11:159.
60. Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. *Stat Med* 2013;32(6):914-30.
61. Debray TP, Schuit E, Efthimiou O, Reitsma JB, Ioannidis JP, Salanti G, Moons KG, GetReal W. An overview of methods for network meta-analysis using individual participant data: when do benefits arise? *Stat Methods Med Res* 2016.
62. 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(8):857-64.
63. Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS One* 2014;9(1):e86754.
64. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making* 2013;33(5):641-56.

65. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010;29(7-8):932-44.
66. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Method* 2012;3:98-110.
67. Law M, Jackson D, Turner R, Rhodes K, Viechtbauer W. Two new methods to fit models for network meta-analysis with random inconsistency effects. *BMC Med Res Methodol* 2016;16:87.
68. Jackson D, Law M, Barrett JK, Turner R, Higgins JP, Salanti G, White IR. Extending DerSimonian and Laird's methodology to perform network meta-analyses with random inconsistency effects. *Stat Med* 2016;35(6):819-39.
69. Jackson D, Barrett JK, Rice S, White IR, Higgins JP. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Stat Med* 2014;33(21):3639-54.
70. Veroniki AA, Mavridis D, Higgins JP, Salanti G. Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study. *BMC Medical Research Methodology* 2014;14(1):106.
71. Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bagos P, Salanti G. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Stat Med* 2014;33(13):2275-87.
72. Efthimiou O, Mavridis D, Riley RD, Cipriani A, Salanti G. Joint synthesis of multiple correlated outcomes in networks of interventions. *Biostatistics* 2015;16(1):84-97.
73. Hong H, Carlin BP, Chu H, Shamliyan TA, Wang S, Kane RL. A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons. Rockville (MD), 2013; Agency for Healthcare Research and Quality (US); Report No.: 13-EHC004-EF
74. Hong H, Carlin BP, Shamliyan TA, Wyman JF, Ramakrishnan R, Sainfort F, Kane RL. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Med Decis Making* 2013;33(5):702-14.
75. Hong H, Chu H, Zhang J, Carlin BP. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods* 2016;7(1):6-22.
76. Jackson D, Bujkiewicz S, Law M, White IR, Riley RD. A matrix-based method of moments for fitting multivariate network meta-analysis models with multiple outcomes and random inconsistency effects. *Biometrics (in-press)* 2017.
77. Del Giovane C, Vacchi L, Mavridis D, Filippini G, Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. *Stat Med* 2013;32(1):25-39.
78. Owen RK, Tincello DG, Abrams KR. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. *Value Health* 2015;18(1):116-26.
79. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
80. Thom HH, Capkun G, Cerulli A, Nixon RM, Howard LS. Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an application to pulmonary arterial hypertension. *BMC Med Res Methodol* 2015;15:34.
81. Saramago P, Woods B, Weatherly H, Manca A, Sculpher M, Khan K, Vickers AJ, MacPherson H. Methods for network meta-analysis of continuous outcomes using individual patient data: a case study in acupuncture for chronic pain. *BMC Med Res Methodol* 2016;16(1):131.
82. Saramago P, Chuang LH, Soares MO. Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data. *BMC Med Res Methodol* 2014;14:105.
83. Jansen JP. Network meta-analysis of individual and aggregate level data. *Res Synth Methods* 2012;3(2):177-90.
84. Hong H, Fu H, Price KL, Carlin BP. Incorporation of individual-patient data in network meta-analysis for multiple continuous endpoints, with application to diabetes treatment. *Stat Med* 2015;34(20):2794-819.

85. Dagne GA, Brown CH, Howe G, Kellam SG, Liu L. Testing moderation in network meta-analysis with individual participant data. *Stat Med* 2016;35(15):2485-502.
86. Jenkins D, Czachorowski M, Bujkiewicz S, Dequen P, Jonsson P, Abrams KR. Evaluation of Methods for the Inclusion of Real World Evidence in Network Meta-Analysis - A Case Study in Multiple Sclerosis. *Value Health* 2014;17(7):A576.
87. Efthimiou O, Mavridis D, Debray TP, Samara M, Belger M, Siontis GC, Leucht S, Salanti G, GetReal Work P. Combining randomized and non-randomized evidence in network meta-analysis. *Stat Med* 2017; 36: 1210–1226.
88. Crequit P, Trinquart L, Yavchitz A, Ravaud P. Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. *BMC medicine* 2016;14:8.
89. Vale CL, Burdett S, Rydzewska LH, Albiges L, Clarke NW, Fisher D, Fizazi K, Gravis G, James ND, Mason MD, Parmar MK, Sweeney CJ, Sydes MR, Tombal B, Tierney JF, Group STS. Addition of docetaxel or bisphosphonates to standard of care in men with localised or metastatic, hormone-sensitive prostate cancer: a systematic review and meta-analyses of aggregate data. *Lancet Oncol* 2016;17(2):243-56.
90. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.
91. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JP, Straus S, Thorlund K, Jansen JP, Mulrow C, Catala-Lopez F, Gotzsche PC, Dickersin K, Boutron I, Altman DG, Moher D. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015;162(11):777-84.
92. Riley RD, Ensor J, Jackson D, Burke DL. Deriving percentage study weights in multi-parameter meta-analysis models: with application to meta-regression, network meta-analysis, and one-stage IPD models. *Statistical Methods in Medical Research* 2017 (in-press).
93. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy* 2012;17(1):1-2.
94. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol* 2014;67(2):138-43.
95. Dias S, Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. *Res Synth Methods* 2016;7(1):23-8.
96. Bafeta A, Trinquart L, Seror R, Ravaud P. Reporting of results from network meta-analyses: methodological systematic review. *BMJ* 2014;348:g1741.
97. Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood AJ, International Stroke Trial Collaborative G. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9(26):1-134, iii-iv.
98. Jackson D, Riley RD. A refined method for multivariate meta-analysis and meta-regression. *Stat Med* 2014;33(33):541-54.
99. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326(7387):472.
100. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009;338:b1147.
101. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How Vague is Vague? A simulation study of the impact of the use of vague prior distributions in MCMC. *Stat Med* 2005;24:2401-28.
102. Burke DL, Bujkiewicz S, Riley RD. Bayesian bivariate meta-analysis of correlated effects: Impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences. *Stat Methods Med Res* 2017 (in-press).
103. Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: the network graphs package. *Stata Journal* 2015;15:905-50.

Figure 1: Publication of network meta-analysis articles over time

(a): Applied articles reporting a systematic reviews using network meta-analysis to compare at least four treatments published between 2005 and 2014. as assessed by Petropoulou et al. [9]



* 6 were also published before 2005, and 43 were published in 2015 up to April.

(b): Methodological articles, tutorials, and articles with empirical evaluation of methods for network meta-analysis published between 2005 and 2014 (as assessed by Efthimiou et al.[10] and available from www.zotero.org/groups/wp4_-_network_meta-analysis)

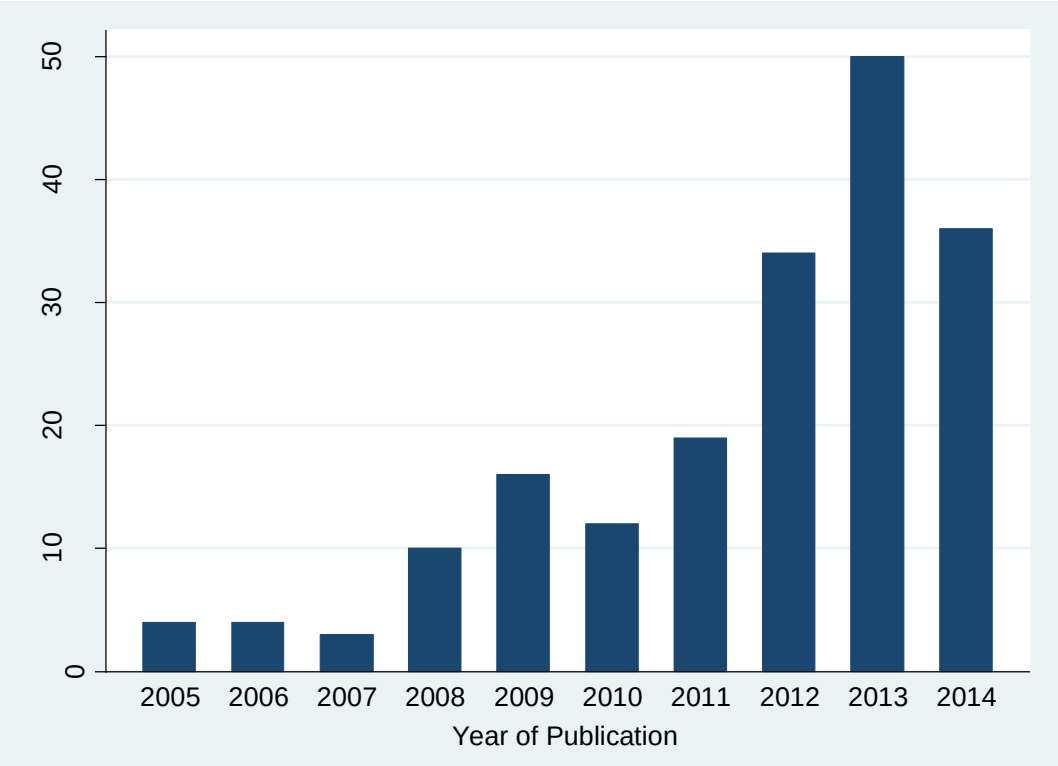
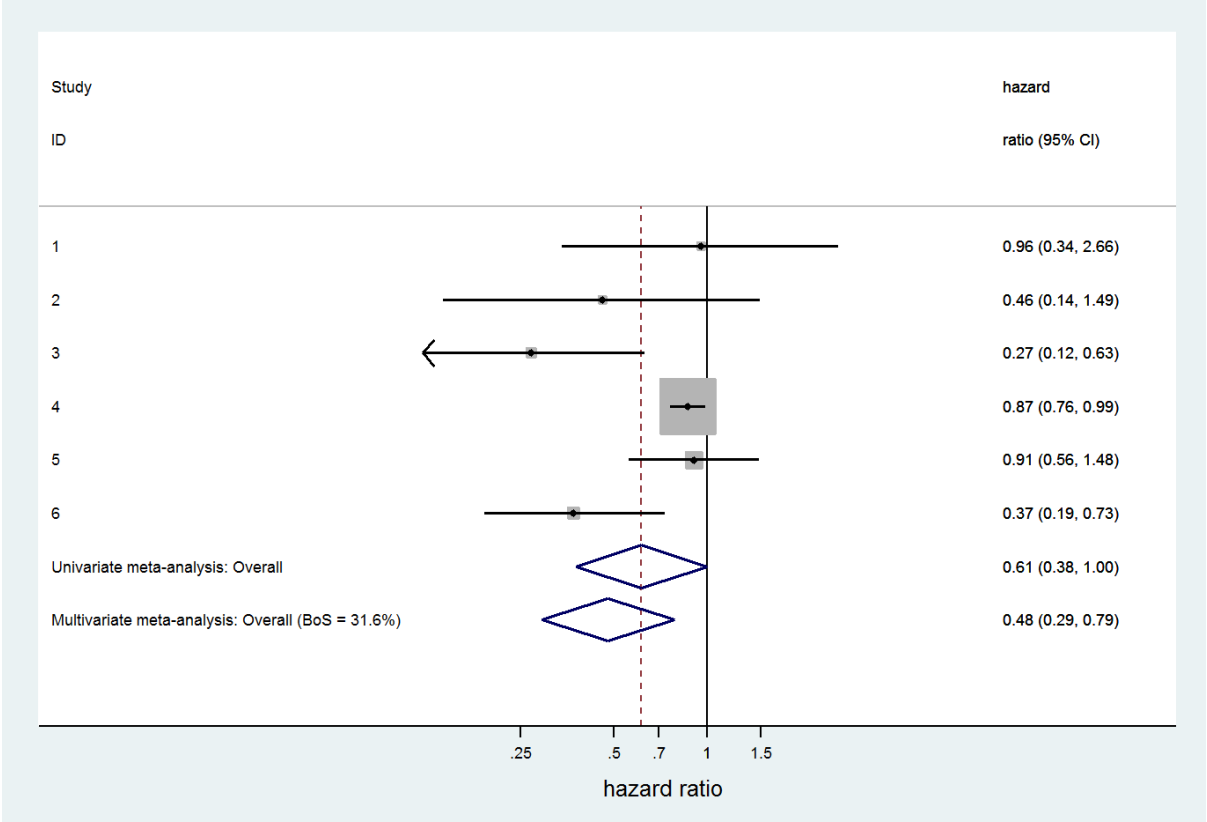
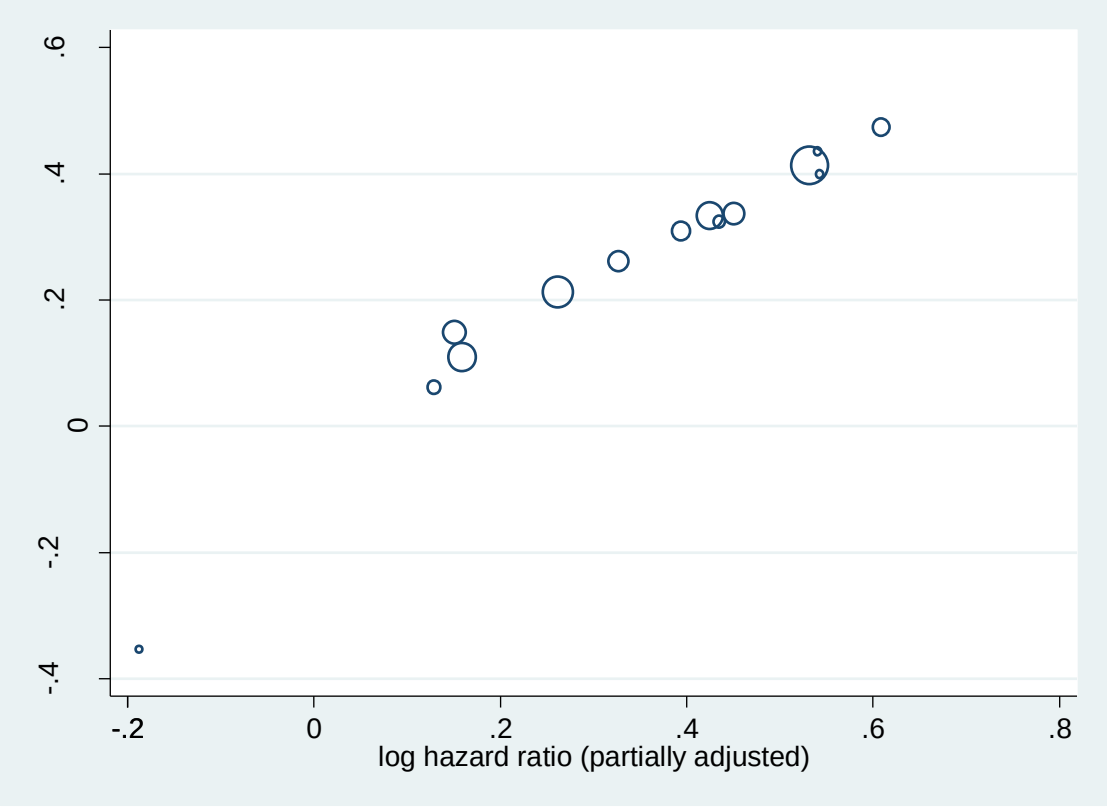


Figure 2: Forest plot for the prognostic effect of progesterone on cancer-specific survival (CSS) in endometrial cancer, with summary results for univariate and multivariate meta-analysis



N.B. The multivariate meta-analysis of CSS and progression-free survival (PFS) used the approach of Riley et al. to handle missing within-study correlations, via restricted maximum likelihood (REML) estimation.[44] Heterogeneity was similar in both univariate and multivariate meta-analyses (I^2 = 70%).

Figure 3: Strong observed correlation (linear association) between the log hazard ratio estimates of the partially and 'fully' adjusted effect of fibrinogen on the rate of cardiovascular disease



NB. The size of each circle is proportional to the precision (inverse of the variance) of the 'fully' adjusted log hazard ratio estimate (i.e. larger circles indicate more precise study estimates). Hazard ratios were derived in each study separately from a Cox regression, indicating the effect of a 1 g/L increase in fibrinogen on the rate of CVD

Figure 4: Visual representation of direct and indirect evidence toward the comparison of A versus B. (adapted from Song et al.[58])

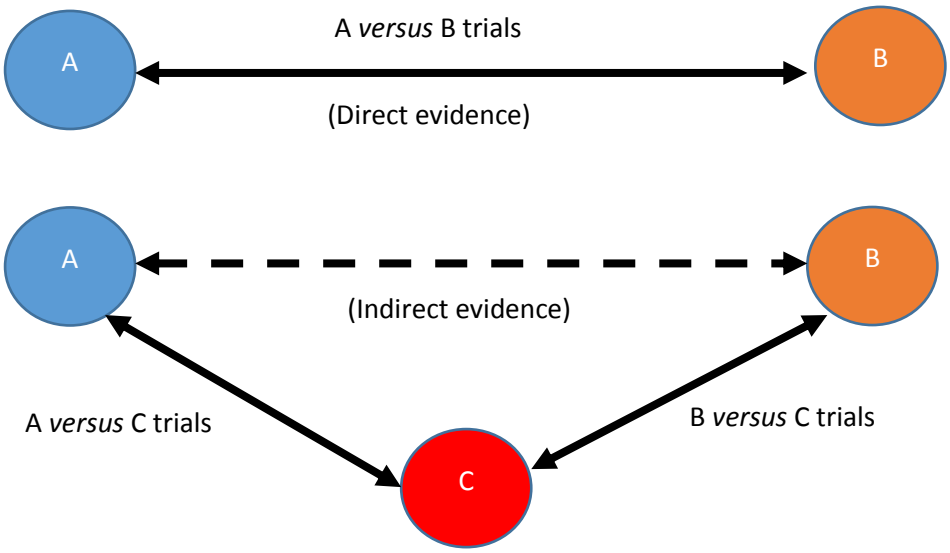
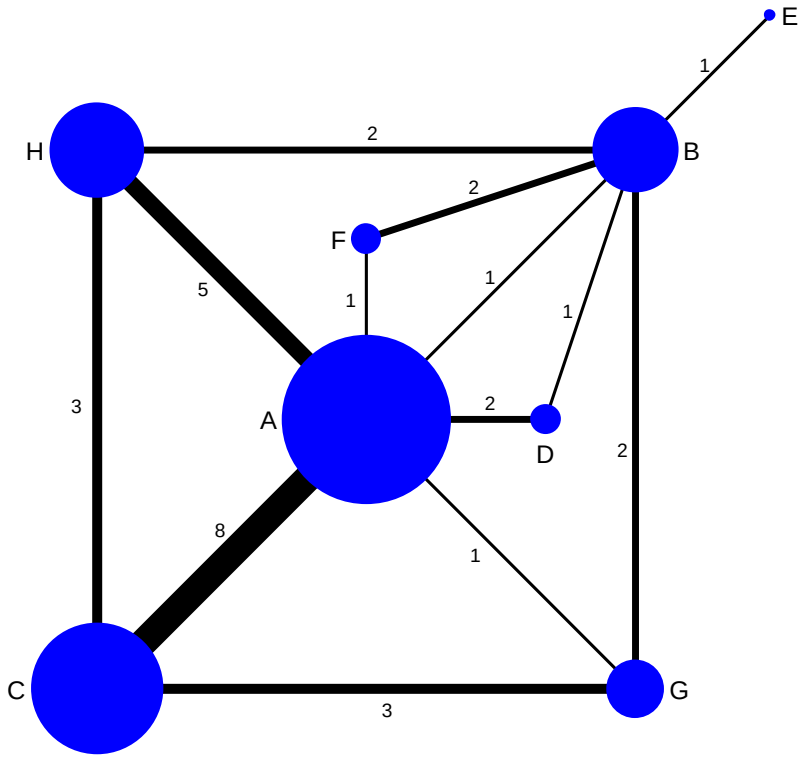


Figure 5: Network map of the direct comparisons available in the 28 trials examining the effect of eight thrombolytics (labelled A to H) on 30-35 days mortality in patients with acute myocardial infarction.



NB Each node (circle) represents a different treatment, and its size is proportional to the number of trials it is directly examined in. The width of the line joining two nodes is proportional to the number of trials that directly compare the two respective treatments (the number is also shown next to the line). Where no line directly joins two nodes (e.g. C and D), this indicates there was no trial that directly compared the two respective treatments.

A = Streptokinase;

B = Accelerated altepase;

C = Alteplase;

D = Streptokinase + alteplase;

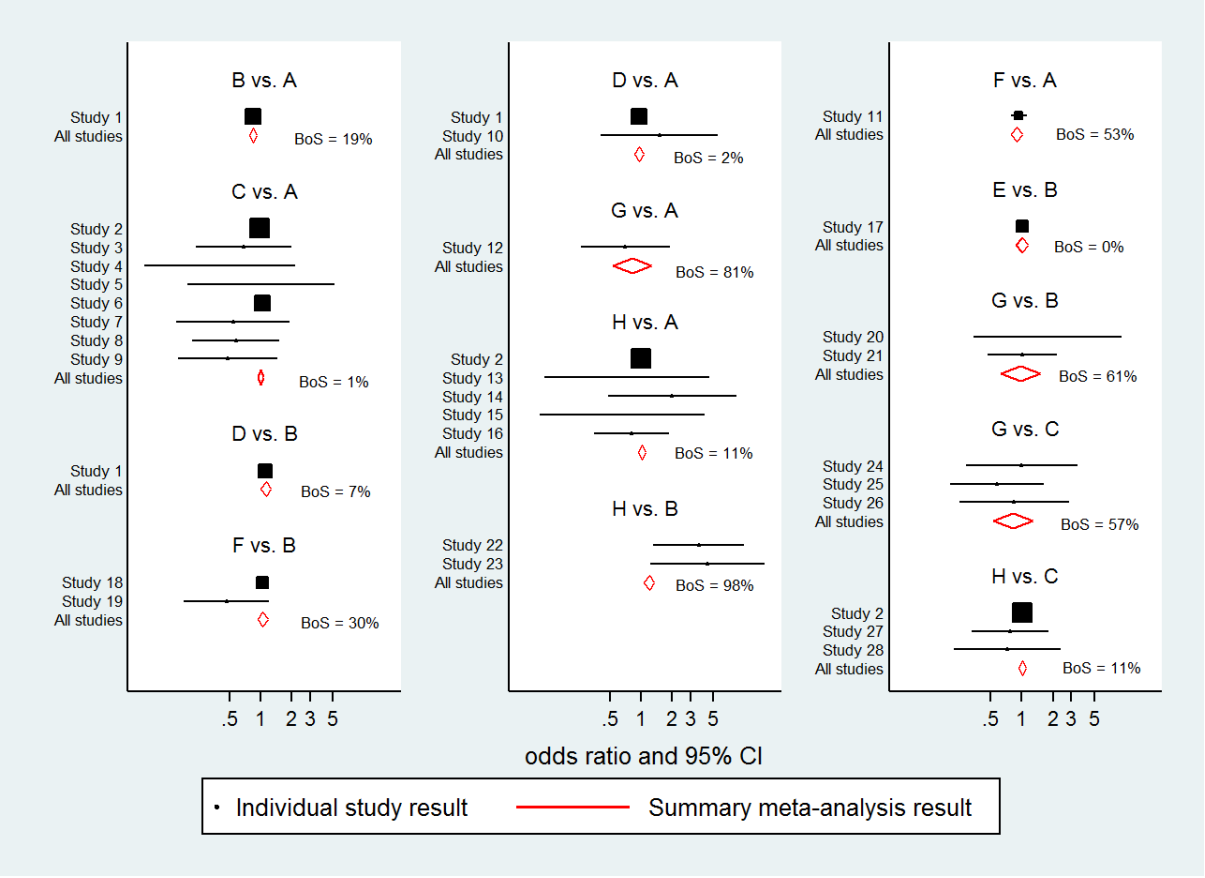
E = Tenecteplase;

F = Reteplase;

G = Urokinase;

H = Anti-streptilase

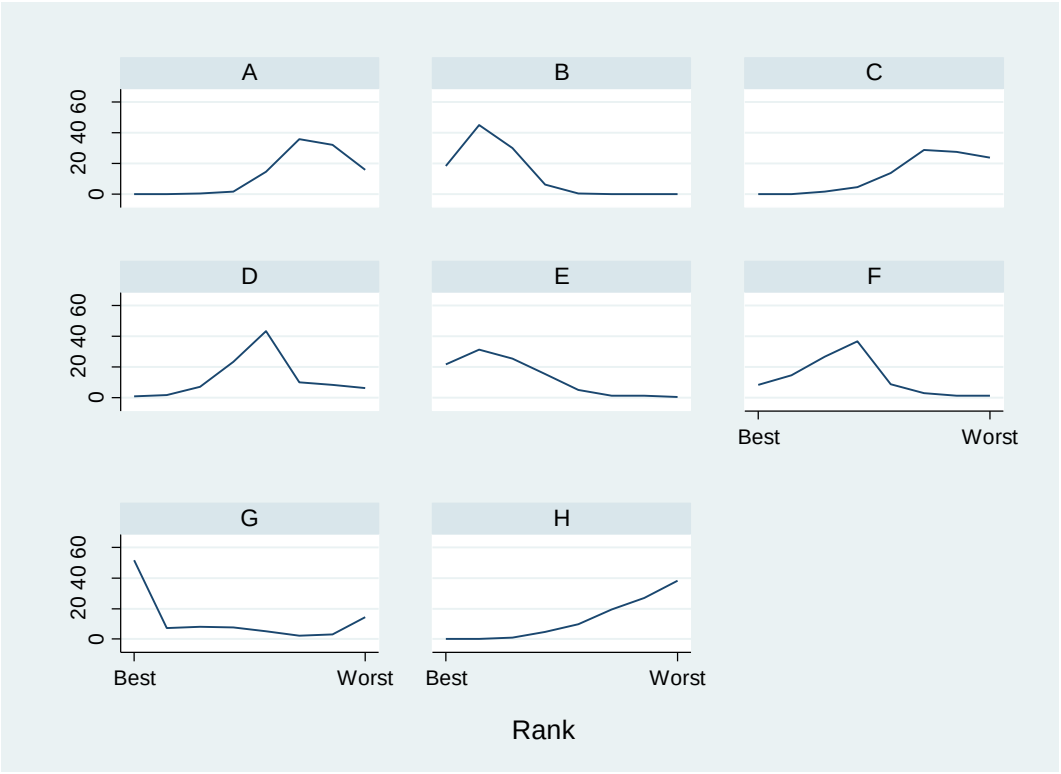
Figure 6: Extended forest plot showing the network meta-analysis results for all comparisons where direct evidence was available in at least one trial



Each square denotes the odds ratio estimate for that study, with the size of the square proportional to the number of patients in that study, and the corresponding horizontal line denotes the confidence interval. The centre of each diamond denotes the summary odds ratio from the network meta-analysis, and the width of the diamond provides its 95% confidence interval. 'BoS' denotes the borrowing of strength statistic, which can range from 0% to 100%.

Figure 7: Plots of the ranking probability for each treatment considered in the thrombolytics network meta-analysis

(a) the probability scale



(b) the cumulative probability scale

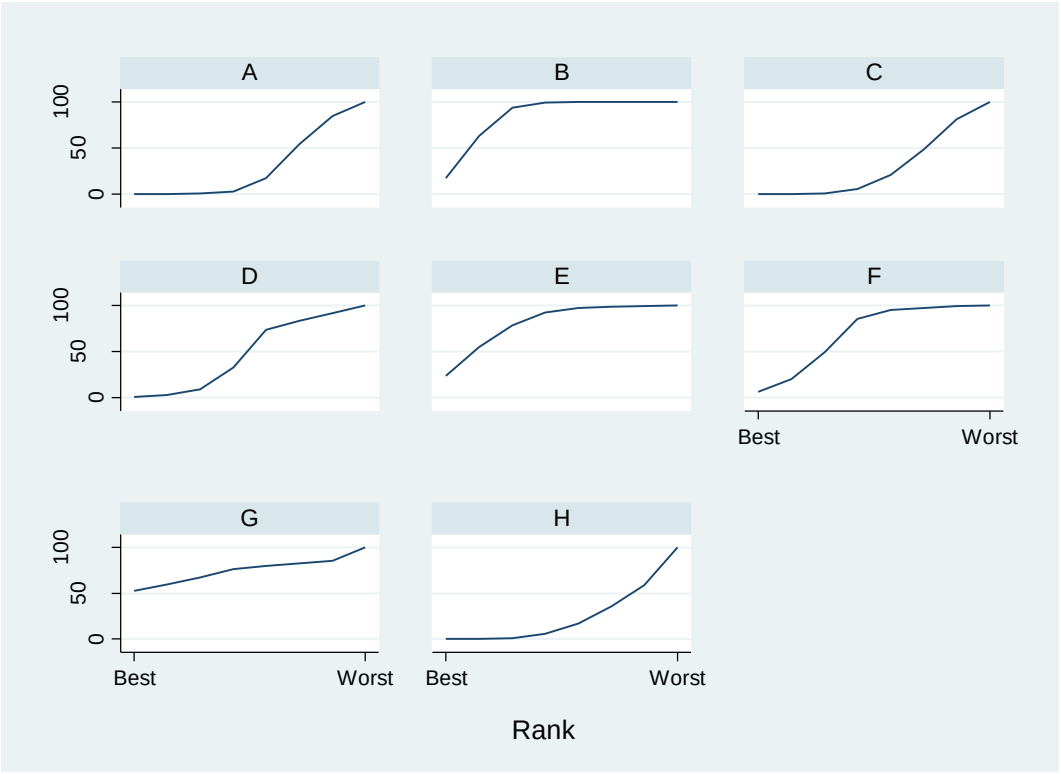


Table 1: Summary of the multivariate and network meta-analysis approaches

	Multivariate meta-analysis of multiple outcomes	**Network meta-analysis of multiple treatment comparisons**
What is the context?	Primary research studies report different outcomes, and thus a separate meta-analysis for each outcome will utilise different studies	Randomised trials evaluate different sets of treatments, and thus a separate (pair-wise) meta-analysis for each treatment comparison (contrast) will utilise different studies
What is the rationale for the method?	• To allow all outcomes and studies to be jointly synthesised in a single meta-analysis model • To account for the correlation amongst outcomes to gain more information	• To enable all treatments and studies to be jointly synthesised in a single meta-analysis model • To allow Indirect evidence (e.g. about A versus B from trials of A versus C and B versus C) to be incorporated
What are the benefits of the method?	• Accounting for correlation enables each outcome's meta-analysis result to utilise the data for *all* outcomes • This usually leads to more precise conclusions (narrower confidence intervals) • It may reduce the impact of selective outcome reporting	• It provides a coherent meta-analysis framework for summarising and comparing (ranking) the effects of all treatments simultaneously • The incorporation of Indirect evidence often leads to substantially more precise summary results (narrower confidence intervals) for each treatment comparison
When should the method be considered?	• When multiple correlated outcomes are of interest, with large correlation amongst them (e.g. > 0.5 or < -0.5) and a high percentage of trials with missing outcomes; or • When a formal comparison of the effects on different outcomes is needed	• When a formal comparison of the effects of multiple treatments is required • When recommendations are needed about the best (or few best) treatments
What are the potential pitfalls of the method?	• Obtaining and estimating within-study and between-study correlations is often difficult • The information gained by utilising correlation is often small and may not change clinical conclusions • The method assumes outcomes are missing at random, which may not hold when there is selective outcome reporting • Simplifying assumptions may be needed to deal with a large number of unknown variance parameters	• Indirect evidence arises via a consistency assumption; i.e. the *relative* effects of three or more treatments observed directly in some trials is (on average) the same in other trials where they are unobserved. This assumption should be checked but there is usually low power to detect inconsistency • Ranking treatments can be misleading due to imprecise summary results, e.g. a treatment ranked first may also have a high probability of being ranked last • Simplifying assumptions may be needed to deal with a large number of unknown variance parameters