Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

The statistical importance of a study for a network meta-analysis estimate

Gerta Rücker (ruecker@imbi.uni-freiburg.de)
Universitatsklinikum Freiburg https://orcid.org/0000-0002-2192-2560

Adriani Nikolakopoulou
Institute of Social and Preventive Medicine (ISPM), University of Bern

Theodore Papakonstantinou
Institute of Social and Preventive Medicine (ISPM), University of Bern

Georgia Salanti
Institute of Social and Preventive Medicine, University of Bern

Richard D. Riley
Centre for Prognosis Research, Rsearch Institute for Primary care and Health Sciences, Keele University

Guido Schwarzer
Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg

Research article

Keywords: Network meta-analysis, Study weight, Study contribution, Study importance

Posted Date: October 1st, 2019

DOI: https://doi.org/10.21203/rs.2.15672/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published on July 16th, 2020. See the published version at https://doi.org/10.1186/s12874-020-01075-y.

RESEARCH

The statistical importance of a study for a network meta-analysis estimate

Gerta Rücker[1*], Adriani Nikolakopoulou[2], Theodore Papakonstantinou[2], Georgia Salanti[2], Richard D Riley[3] and Guido Schwarzer[1]

[*]Correspondence: ruecker@imbi.uni-freiburg.de
[1]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Stefan-Meier-Strasse 26, 79104 Freiburg, Germany
Full list of author information is available at the end of the article

Abstract

Background: In pairwise meta-analysis, the contribution of each study to the pooled estimate is given by its weight, which is defined based on the inverse variance of the estimate from that study. For network meta-analysis (NMA), the contribution of direct (and indirect) evidence is easily obtained from the diagonal elements of the hat matrix. It is, however, not fully clear how to generalize this to the percentage contribution of each study to a NMA estimate.

Methods: After briefly discussing available approaches, we want to question whether it is possible to obtain unique percentage contributions and discuss another approach. We define the importance of each study for a NMA estimate by the reduction of the estimate's variance when adding the given study to the others. An equivalent interpretation is the relative loss in precision when the study is left out.

Results: Importances are values between 0 and 1. An importance of 1 means that the study is an essential link of the pathway in the network connecting one of the treatments with the other. These numbers in general do not add to one and thus cannot be interpreted as 'percentage contributions'.

Conclusions: Importances generalize the concept of weights in pairwise meta-analysis in a natural way. Moreover, they are uniquely defined, easily calculated, and have an intuitive interpretation. We give some real examples for illustration.

Keywords: Network meta-analysis; Study weight; Study contribution; Study importance

Background

Pairwise meta-analysis (pairwise MA) is used to summarize the evidence for a treatment effect from all eligible studies that compared the two interventions of interest. Network meta-analysis (NMA) extends this approach to an arbitrary number of interventions. In two-stage pairwise MA, the contribution of each study to the pooled estimate is measured by its weight, which depends on the type of data, the chosen summary measure, and the chosen statistical model. For example, for mean differences usually a study's weight is the inverse of the variance of the estimated mean difference for that study.

To describe the influence of a study on a MA estimate, we distinguish between approaches that focus on the study's *contribution* in the sense of attributing a 'weight' to each study, mainly depending on the variances, and approaches that also take into account the direction and size of each treatment effect, and which are

seeking for 'influential' or explicitly 'outlying' studies. For the latter, see [1, 2, 3]. In this paper, we concentrate on the first approach; that is, we are mainly interested in measuring study contributions without looking at their effect estimates.

For NMA, while several methods exist to obtain the contribution of direct (and indirect) evidence of each comparison to its own NMA estimate, it is far less obvious how to define the contribution, or the importance, of each study to any (other) effect estimate for a given treatment comparison. Several proposals exist in the literature, based on different approaches. Most of them come with some limitations; their results also do not in general agree [4, 5, 6, 7].

The proportions of direct and indirect evidence have been investigated in the past. The method of 'back-calculation', which we will describe in the methods section, goes back to the Bucher method [8] and was proposed by Dias and others in a Bayesian framework [9]. It was also given in a frequentist context [10]. In frequentist NMA, the direct evidence proportion of a study or a comparison is easily obtained from the diagonal elements of the respective hat matrix [10]. As an alternative, Dias and others suggested 'node splitting', which means estimating the indirect evidence for a comparison by modeling out all studies that provide direct information for this comparison [9]. This method was developed further [11] and called 'side splitting' by others [12]. Whereas White [12] interpreted the term 'side' as an edge in the network graph, others coined it SIDE as an abbreviation of 'Separating Indirect and Direct Evidence' [13, 14]. Noma and others proposed quantifying the indirect evidence based on a factorization of the total likelihood into separate component likelihoods [15]. So far, these authors did not undertake to define or estimate the contribution of each study to a given comparison in the network.

A proposal to this aim, based on the off-diagonal elements of the hat matrix, was made by Salanti and others [4]. Implicitly, this idea also underlies the net heat plot that was suggested by Krahn and others, a visualization where the absolute size of the hat matrix elements are represented by gray squares [16]. However, as an approach to define percentage study contributions, it is associated with problems, as the hat matrix elements are signed and do not add to 1. Papakonstantinou and others, acknowledging this limitation, developed a different concept likewise based on the hat matrix, using ideas by König and others [10], and successfully avoided these deficiencies [7]. Although the proposed algorithm is not strictly deterministic, it was shown empirically that this does not materially impact on the estimated percentage contributions.

Another approach to percentage contributions was published by Jackson and others [5] and also used elsewhere [6, 17]. They suggest decomposing the total variance matrix of parameter estimates from the meta-analysis into study-specific contributions via a decomposition of the observed Fisher information matrix into independent study-specific contributions, which sum up to the total variance matrix. Their diagonal elements can be used to derive percentage study weights for each parameter. They reveal how the variance of a parameter is changed by the inclusion of a particular study, assuming that all variance estimates (within and between-study variances) are fixed at the same value as in the full analysis of all studies. Although this approach is adequate when within-study information is pooled across studies, the study-specific contributions can become distorted in situations where

across-study information contributes to the parameter estimates, as in a network meta-analysis [7].

In the present paper we suggest an approach that deviates from the idea of having 'percentage contributions' to each network estimate summing up to 100%. The structure of the article is as follows. In the methods section we introduce our notation, define direct and indirect evidence proportion based on the hat matrix, and give several interpretations of the quantity 'direct evidence proportion'. We will show that starting from different interpretations of this quantity leads to different generalizations, one of which we will then focus on. In the next section we present two real data sets for illustrating our method. In the results section we illustrate our concept by first applying it to simple special cases and standard networks like pairwise MA, a chain of treatments, and a circle, and then apply it to the two real data examples. We will discuss strengths and limitations of our approach in the discussion section, and the paper ends with a recommendation in the last section (Conclusion).

Methods

This paper is based on the two-stage approach to network meta-analysis. We assume to have estimates of treatment effect from each study, which are then to be combined in a network meta-analysis. We do not deal with one-stage modelling, e.g., of binary data directly based on the binomial distribution.

After presenting our notation and formulas for variance estimates in NMA in the next subsection, we will give definitions for the direct and indirect evidence proportion and also show that several interpretations exist for these quantities. We will then define the importance of a study for any network comparison in the framework of the common effect model (traditionally termed 'fixed effect model') and for the random effects model. While the common effect model assumes that for each comparison all studies in the network are estimating the same (comparison-specific) true effect, the random effects model assumes that the underlying effects of each comparison follow a distribution. Often a normal distribution is assumed [18].

The variance of direct, indirect and network effect estimates in the common effect model

Consider a network meta-analysis (NMA) and any particular comparison c in the network. First, we concentrate on the common effect model. Let $y_{nma}(c)$ be the NMA treatment effect estimate, $y_{dir}(c)$ the direct effect estimate and $y_{ind}(c)$ the indirect effect estimate for c, all based on the common effect model. Direct and indirect paths and thus effects can be assumed as independent. We will treat the random effects model in another subsection below.

Let us write $V_{nma}(c) = \mathrm{Var}(y_{nma}(c)), V_{dir}(c) = \mathrm{Var}(y_{dir}(c)), V_{ind}(c) = \mathrm{Var}(y_{ind}(c))$ for the variance estimates of the various effects and $w_{nma}(c) = [V_{nma}(c)]^{-1}, \ldots$ etc. for the inverse variance weights. Note that, for readability, we omit hats on top of the symbols.

The principle of back-calculation is based on the rules of variance calculation [8, 9]. We have

$$V_{nma}(c) = \left(\frac{1}{V_{dir}(c)} + \frac{1}{V_{ind}(c)}\right)^{-1} = \frac{V_{dir}(c) \cdot V_{ind}(c)}{V_{dir}(c) + V_{ind}(c)}. \tag{1}$$

Solving this for $V_{ind}(c)$ gives

$$V_{ind}(c) = \frac{V_{nma}(c)V_{dir}(c)}{V_{dir}(c) - V_{nma}(c)}$$

which leads to

$$\frac{V_{nma}(c)}{V_{ind}(c)} = \frac{V_{dir}(c) - V_{nma}(c)}{V_{dir}(c)} = 1 - \frac{V_{nma}(c)}{V_{dir}(c)} \tag{2}$$

$$\frac{V_{nma}(c)}{V_{dir}(c)} = 1 - \frac{V_{nma}(c)}{V_{ind}(c)} = \frac{V_{ind}(c) - V_{nma}(c)}{V_{ind}(c)}. \tag{3}$$

Expressing the right hand side of equation (3) using inverse variance weights, we obtain (weights denoted by w, indexed accordingly)

$$\frac{V_{nma}(c)}{V_{dir}(c)} = 1 - \frac{w_{ind}(c)}{w_{nma}(c)} = \frac{w_{nma}(c) - w_{ind}(c)}{w_{nma}(c)}. \tag{4}$$

Direct and indirect evidence proportion

The direct evidence proportion of comparison c can be defined via the inverse variance weights as

$$p(c) := \frac{w_{dir}(c)}{w_{dir}(c) + w_{ind}(c)} = \frac{[V_{dir}(c)]^{-1}}{[V_{dir}(c)]^{-1} + [V_{ind}(c)]^{-1}} = \frac{V_{ind}(c)}{V_{dir}(c) + V_{ind}(c)}. \tag{5}$$

Inserting (1) into (5) and using also (2) and (3), we obtain for the proportions of direct and indirect evidence

$$p(c) = \frac{V_{nma}(c)}{V_{dir}(c)}; \quad 1 - p(c) = \frac{V_{nma}(c)}{V_{ind}(c)}. \tag{6}$$

In practice, the values of $p(c)$ are found on the diagonal of the hat matrix; users of the R package **netmeta** obtain them via function `netmeasures()` [19]. We now show that different interpretations for $p(c)$ are possible and that these lead to different concepts of generalization.

Interpretation 1: Proportion of direct evidence

The first interpretation of $p(c)$, an immediate consequence of the definition (5), is that it describes the proportion of network precision for comparison c attributed to direct evidence (from pairwise MA), in short, the contribution of direct evidence to the network estimate c. Accordingly, $1 - p(c)$ represents the contribution of indirect evidence to this estimate.

Interpretation 2: Reduction of the variance of a direct comparison by adding indirect information

Equation (6) provides another interpretation: $p(c)$ is the proportion to which the variance of a pairwise MA is shrunk when indirect evidence from the whole network is added to the direct evidence, or, in other words, when all network information is used.

Interpretation 3: Relative reduction of the variance of a comparison based solely on indirect evidence when adding direct information

This interpretation is suggested by equation (3): $p(c)$ is the relative reduction of the variance of a comparison with only indirect evidence when information from the direct comparison is added. Similarly, $p(c)$ could be interpreted as the loss in precision of a comparison when removing direct evidence (i.e., using exclusively indirect evidence). This interpretation is motivated by equation (4).

Whatever interpretation is preferred, high values mean a high importance of the direct comparison for itself, and low values mean low importance. We note that the concept, as described, is limited to a given comparison, searching to quantify the contribution of the direct evidence from this comparison to its own NMA estimate. We want to generalize the concept of 'importance for its own NMA estimate' to a concept of the importance of a comparison/a study for the NMA estimate of *any (other)* comparison.

The idea of comparing variances also underlies the 'Borrowing of strength' (BoS) measure, developed in the more general framework of multivariate meta-analysis [5, 6]. With respect to interpretation, Copas and others distinguish between 'Direct interpretation' (which corresponds to our interpretation 1), 'Add-one-in interpretation' and 'Leave-one-out interpretation' (which both refer to our interpretation 3) [17] .

Different interpretations suggest different ideas of generalizing the $p(c)$. A generalization of interpretation 1 (i.e., splitting direct and indirect contribution) aims to determine the contributions of each comparison (or, alternatively, of each study) to a given NMA estimate, such that these contributions add up to 100% ('percentage contributions'). This means splitting $1 - p(c)$, the proportion of indirect evidence, further into parts coming from different comparisons or studies, as done by [7].

Starting from interpretation 2 would mean looking for a quantity that describes the proportion to which the variance of a given direct estimate at hand decreases by adding (indirect) evidence from a particular other study. However, adding a single study to a given direct comparison makes only sense if the enlarged network is connected, that is if the new study and the comparison in question have treatments in common or if further studies are added. Thus, interpretation 2 does not seem to be a good starting point for generalization.

Therefore we focus on a generalization that is motivated by interpretation 3.

Variance reduction by adding direct information: the importance of each study

First concentrating on the common effect model, we ask for the amount by which the variance of an estimate from only indirect evidence is reduced if direct information is added, or the relative loss in precision when direct evidence is removed. Note that

there is no reason to assume that these quantities add up to 100%. We will come back to this point later.

The *importance of each study* for each NMA estimate is now defined as follows:

Conduct a NMA for the given network, called NMA_{all}, and then repeat the following steps for each study i in turn:

1 Set the standard error of the treatment effect estimates for all direct comparisons from study i to a very large value (e.g., 10000), thus downweighting study i to practically zero.

2 Conduct a NMA for the network with downweighted study i. Let us denote this result NMA_{-i}.

3 For all comparisons c, define the importance of study i for comparison c as

$$p(i,c) = 1 - \frac{V_{\text{NMA}_{all}}(c)}{V_{\text{NMA}_{-i}}(c)} = \frac{V_{\text{NMA}_{-i}}(c) - V_{\text{NMA}_{all}}(c)}{V_{\text{NMA}_{-i}}(c)}.$$

The first step is intended to 'remove' study i from the network. Technically, however, instead of removing study i, which could lead to a disconnected network (rendering the calculation of $V_{\text{NMA}_{-i}}(c)$ impossible), we set its variance to a very large value. This data augmentation approach goes back to White and others [20] and was also used by Riley and others [6]. The third step is motivated by equation (3). We have $0 \leq p(i,c) \leq 1$ for all studies i and comparisons c. We emphasize, however, that it makes no sense to add up these importances across all studies, as they do not sum up to 100% (in fact, the sum is often larger). We do not call them (percentage) contributions. Rather, they measure the importance of a study for a comparison.

The idea can be illustrated by comparing a network of studies to a network of roads in a town. We consider the traffic from some place A (a node in the network) to another place B. The precision (or the weight) can be interpreted as the transport capacity of the road network between A and B, comparable to the conductance in an electrical network [21]. If a particular road is closed due to construction works, many capacities decrease because some people have to make a detour to go from A to B and thus add to the traffic on other roads. Interpretation 3 refers to the relative reduction of the capacity of the network due to the road closure, which represents the *importance* of this road for the way from A to B.

The algorithm to calculate importances is implemented in R function `netimpact()` in the R package **netmeta** [19, 22], with a data set about Parkinson's disease as example. R code for all examples can be found in Additional File 1, with the resulting plots shown in Additional File 2.

Extension to the random effects model

So far, we used the common effect model to derive the importances. The estimate $\hat{\tau}^2$ of the variance of the random effects did not enter the calculations. Leaving out a study from the data is expected to change $\hat{\tau}^2$. Particularly, the variance may *decrease* if the omitted study contributed a lot to the between-study heterogeneity or inconsistency, resulting in a negative importance for this comparison. A possible workaround is to insert the estimate $\hat{\tau}^2$ from the original network as a common heterogeneity variance estimate for all subnetworks [5]. For pairwise MA this leads to the usual random effects weights, as we will see in the next subsection.

Pairwise meta-analysis

We consider a pairwise MA with inverse variance weighting, such that the (unstandardized) weight of study i is given by $w_i = 1/V_i$ where V_i is the variance of study i. The variance of the pooled common effect estimate is then estimated by $1/\sum_j w_j$. Removing study i from the MA gives another pooled estimate with variance $1/\sum_{j\neq i} w_j$. Equation (4) provides the importance of study i for the pooled estimate

$$\frac{\sum_j w_j - \sum_{j\neq i} w_j}{\sum_j w_j} = \frac{w_i}{\sum_j w_j}$$

which is the relative weight of study i in line with what we would expect.

Note that this only holds if inverse variance weighting is used (e.g., for mean differences or Peto odds ratios), but not in general. For example, due to the different weighting method, it is not true for binary outcomes when using the Mantel-Haenszel method. Nevertheless, the importance concept allows to derive approximate 'study weights' for one-stage pairwise meta-analyses based on a generalized linear model, such as logistic regression. We provide an example in Additional File 1.

For the random effects model, we use an estimate of the heterogeneity variance $\hat{\tau}^2$ for the full pairwise MA, and then remove one study in turn while fixing the heterogeneity variance to this value. The same argumentation as above shows that the procedure leads to the usual random effects weights, $w_i^* = 1/(V_i + \hat{\tau}^2)$ and corresponding relative weights.

Data sets

Parkinson data

This network consists of seven studies comparing five treatments (placebo, coded 1, and four active drugs coded 2 to 5). The outcome is the mean lost work-time reduction in patients given dopamine agonists as adjunct therapy in Parkinson's disease, given as sample size, mean and standard deviation in each trial arm. The data is used as an example in the supplementary material of [23] and included in the R package **netmeta** [19].

Thrombolytic data

This data set, originally published by Boland and others [24], was presented by Lu and Ades [25] and sucessively analyzed by many others, for example Riley and others [6]. This network consists of 28 studies (13 designs, i.e., different combinations of treatments in a study) of 8 treatments after acute myocardial infarction. We follow Riley and others [6] denoting these treatments A, B, C, D, E, F, G, and H. Figure 1 shows the network graph for the thrombolytic data which are provided in Additional File 3.

Results

We first apply our method to a number of hypothetical examples that nevertheless lead to insight into the interpretation of our new measure of importance.

Hypothetical networks

A chain of $n-1$ studies connecting n treatments

Suppose we have three studies comparing A to B, B to C, and C to D with variances V_1, V_2, V_3. We look at comparison A:D. The direct evidence proportion for comparison A:D is 0, the indirect evidence proportion is 1. The variance of the NMA (i.e., the indirect) estimate for comparison A:D is $V_{\mathrm{NMA}_{all}}(\mathrm{A{:}D}) = V_1 + V_2 + V_3$. If one of the studies, regardless which, is omitted, the variance becomes infinite, and the importance of this study for comparison A:D becomes 1. The interpretation is that this study (like the others) is of maximum importance for the comparison, which is indeed true. For the approach by Papakonstantinou and others, henceforth called contributions approach, each study would contribute 1/3 to the estimate of comparison A:D [7].

A circle of n treatments with equal variances

Suppose we have a closed circular network of n treatments connected by exactly n two-arm studies, each comparing two treatments in turn like in Figure 2 (left, here $n = 7$) and each having variance 1.

We consider an arbitrary comparison $c = t_1 : t_2$ of treatments t_1 and t_2 ($t_1 \neq t_2$) in the network such that the distance (the number of steps) from t_1 to t_2 is $1 \leq l \leq n/2$. The NMA variance of comparison c is

$$V_{\mathrm{NMA}_{all}}(c) = \frac{1}{\frac{1}{l} + \frac{1}{n-l}} = \frac{l(n-l)}{n}.$$

For example, comparison A:D (distance 3) has the NMA variance 3*4/7. If a study is removed from the network, evidence for comparison c can only go one way. If, for example, the study connecting B and C is removed from the network, the variance for comparison A:D becomes 4, because the shorter connection via B and C is broken. In general, if the removed study lies on the shorter path from t_1 to t_2 (length l), the variance becomes $n-l$, if the removed study lies on the longer path, the variance becomes l. Thus the importance of each study i on the shorter path for comparison c is

$$p(i,c) = 1 - \frac{V_{\mathrm{NMA}_{all}}(c)}{V_{\mathrm{NMA}_{-i}}(c)} = 1 - \frac{l(n-l)}{n(n-l)} = 1 - \frac{l}{n} = \frac{n-l}{n}$$

and the importance of each study j on the longer path for comparison c is

$$p(j,c) = 1 - \frac{V_{\mathrm{NMA}_{all}}(c)}{V_{\mathrm{NMA}_{-j}}(c)} = 1 - \frac{l(n-l)}{nl} = 1 - \frac{n-l}{n} = \frac{l}{n}.$$

Thus, plausibly, each of the l studies on the shorter path has greater importance for the comparison than the $n-l$ studies on the longer path, as we have presumed $l \leq n-l$. For the example, the importance of study B:C for comparison A:D is (7 - 3)/7 = 4/7. Particularly, it follows that the direct evidence proportion for each pair that is directly compared (i.e., adjacent, $l = 1$) is $(n-1)/n$ and the indirect evidence proportion is $1/n$, while for all other pairs the direct evidence proportion is

0 and the indirect evidence proportion is 1. By contrast, the contributions approach would attribute a contribution of $\frac{n-l}{nl}$ to each piece on the shorter path and $\frac{l}{n(n-l)}$ to each piece on the longer path, such that the sum of all contributions is 1.

A network with a bridge

We consider the network given on the right-hand panel of Figure 2 with seven studies, all again having variance 1. We call comparison C:D a bridge [7, Supplementary file 3 therein]. Table 1 gives the importances of each study for all comparisons. Study CD has importance 1 for all comparisons between the two parts of the network (A:D, A:E, A:F, B:D, B:E, B:F, C:D, C:E, C:F) and importance 0 for all comparisons within the same part of the network (A:B, A:C, B:C, D:E, D:F, E:F). Studies AB, AC and BC have no importance for comparisons outside the triangle ABC (C:D, C:E, C:F, D:E, D:F, E:F), and vice versa for triangle DEF. The table also shows that the direct evidence proportion for comparisons A:B, A:C, B:C, D:E, D:F, and E:F is 2/3, and the direct evidence proportion for comparison C:D (the bridge) is 1.

To compare this with the contributions approach [7], Table 2 shows the contributions of each study to a comparison in the bridge network. For all comparisons of treatments from different parts of the network the values of contributions differ from those of importances. This is because the weights attributed to a study by the contributions approach are the lower, the greater the network distance between the treatments is.

A generic triangle

Consider a NMA with three treatments A, B and C and three studies comparing A to B, A to C, and B to C with variances V_{AB}, V_{AC}, V_{BC}. We focus on comparison A:B. Its direct estimate has variance V_{AB} and its NMA estimate has variance

$$V_{\mathrm{NMA}_{all}}(A:B) = \frac{1}{V_{AB}^{-1} + (V_{AC} + V_{BC})^{-1}} = \frac{V_{AB}\ (V_{AC} + V_{BC})}{V_{AB} + V_{AC} + V_{BC}}$$

The direct evidence proportion for A:B is

$$\frac{V_{AC} + V_{BC}}{V_{AB} + V_{AC} + V_{BC}}.$$

If one of the studies on the indirect pathway from A to B (say AC) is removed, only direct evidence remains, and we get its importance for comparison A:B as

$$p(AC, A:B) = 1 - \frac{V_{\mathrm{NMA}_{all}}(A:B)}{V_{AB}} = 1 - \frac{V_{AC} + V_{BC}}{V_{AB} + V_{AC} + V_{BC}} = \frac{V_{AB}}{V_{AB} + V_{AC} + V_{BC}},$$

the same if BC is removed. In other words, it does not matter whether we remove AC or BC or which of them has smaller variance, the importance of both studies is equal and also equal to the indirect evidence proportion. The indirect evidence proportion comes from the *combination of studies* AC and BC and depends on the sum of their variances. This example shows that it does not make sense to add up the importances of all studies. It also challenges the idea of breaking up the indirect evidence proportion into additive parts from each study.

Real data networks

We now apply our method to the two real data sets presented before.

Parkinson data

Figure 3 shows the network graph of the Parkinson data [23] (top left panel) and the effect of removing one study after the other in turn (study 1: comparison 1:3; study 2: comparison 1:2; study 3: comparisons 1:2, 1:4, 2:4; studies 4,5: comparison 3:4; studies 6,7: comparison 4:5).

The resulting importances from the common effect model are given in Table 3. The last row of the table provides the direct evidence proportions for each comparison.

We see that, not surprisingly, the most important study is the three-arm study 3. It is important not only for comparisons 1:2, 1:4 and 2:4, but also for comparison 1:3 and the indirect comparisons 1:5, 2:3 and 2:5. However, for comparison 1:2, study 2 is even more important, and for comparison 1:3 study 1 is more important. Study 4 (comparison 3:4) is the most important study for both comparisons 3:4 and 3:5, whereas the less precise Study 5 (likewise comparing 3:4) is less important for all comparisons. For comparison 4:5, only studies 6 and 7 are important. Study 6 is uniformly more important than study 7 (both comparing 4:5).

For each direct comparison we may compare this to the contributions (weights) of each study in a pairwise meta-analysis, given in Table 4. For comparisons solely informed by direct evidence (here comparison 4:5) they agree with the corresponding importances.

Thrombolytic data

Table 5 shows results of our method (random effects model) when applied to the thrombolytic data [6, Tables 3 and 4 therein]. Note that the importance values are not directly comparable to percentage contributions, as they do not add up to 100%. Therefore we find major differences to the percentage contributions given by Riley and others [6]. In all these cases, the importance (as we measure it) exceeds the percentage contribution (as defined in [6]). Particularly, study 17 is the only one including treatment E (it compares B to E), and if it is omitted, E is no more part of the network. Therefore study 17 has maximal importance for comparison A:E, compared to 67.13% contribution following Riley's method. The three-arm study 1 (A:B:D) is more important for comparisons A:E and A:F as when measured by Riley et al.'s % contribution approach. Also study 18, comparing B:F, has higher importance for comparisons A:B, A:E and A:F as when measured by Riley et al.'s % contribution approach. Again, the importance values may be compared to the study weights in pairwise meta-analyses, given in Table 6 for the random effects model.

A visualization as a heatmap is shown in Figure 4. Dark colors mean that a study (in the column) is important for the comparison in the row. It is not surprising that studies 1 and 2, both three-arm studies with over 40.000 patients each, are at the top level of importance for most, but not all comparisons. As mentioned before, study 17 (B:E) is important for all comparisons with E, study 18 (B:F) for all comparisons with F, and study 21 (B:G) for all comparisons with G.

Discussion

In two-stage pairwise MA, the pooled effect estimate is a weighted mean of the study-specific estimates. Relative study weights (for example inverse variance weights) are defined and can be interpreted as proportions or percentages, adding to 100%. Most existing approaches to generalize the concept of 'weights' to NMA aim to define study contributions that can be represented as proportions or percentages, like in pairwise MA. At least, it is possible to quantify the proportion of direct and indirect evidence for each NMA estimate, as outlined in the methods section. NMA estimates are linear combinations of treatment effect estimates from primary studies with coefficients that constitute the rows of the hat matrix. However, these can be negative, and their absolute values do not add to 1. Krahn and others used their absolute values to define the size of the gray squares in the net heat plot [16]. König and others and Papakonstantinou and others interpreted the hat matrix entries as generalized weights [7, 10]. In examples, they show that entries in a hat matrix row can be interpreted as a flow through the network where different signs indicate the direction of the flow. Consequently, it makes little sense to add the values of the entries of the hat matrix, as the direction is embedded.

The concept of importance for the variance

Our concept of study importance does not start from the hat matrix, but interprets the importance of a study to a comparison as the relative reduction of the variance of the estimate when adding the study to the network. We abstain from requesting the values to sum up to 1. For example, a study can be essential for a comparison (like study 17 in the thrombolytic data for all comparisons involving treatment E), thus providing an importance of 1, but other studies may be also (or even equally) important. For instance, in the chain or the circle example, all studies on a path are equally important for the comparison of the path's ends.

Importance and contribution

For pairwise MA there is no ambiguity: the contributions (weights) of all studies add to 1. Likewise, in NMA the direct evidence can be broken down into percentage weights. The division into direct and indirect contribution in NMA (which add to 1) is also possible. However, it is the breakdown of percentage weights for the *indirect* evidence that does not work. Therefore, for the more general situation, we use 'importance' instead of 'contribution', because these two words have different connotation. Coming back to the example of the chain, the importance of each of the studies connecting A to D is 1, meaning that each of these studies is needed for comparing A to D. By contrast, the concept of 'contribution' by Papakonstantinou and others [7] accounts for the fact that though all these studies are *necessary*, none of them alone is *sufficient* for comparing A and D. Therefore for this example they divide 1 by the number of linking studies which is three, leading to a contribution of 1/3 for each study ($1/(n-1)$ in general). However, this approach is not strictly deterministic, as demonstrated in [7, Supplementary File 3 thereof].

Combinations of studies matter

The importance of a study for a comparison must be seen in combination with other studies. Possible extensions could be to define the importance of combinations of

studies, or the importance of a single patient in a study to a NMA estimate. We emphasize that the importance of a study is always conditional on the other studies being included in the network. There is an analogy to a multivariable regression model: If the association of each regressor (covariate) x_i with the dependent variable is considered in isolation, the proportion of explained variance of the dependent variable is given by its coefficient of determination, r_i^2, which is bounded between 0 and 1. When considering more than one covariate, it does not make sense to consider their r_i^2 values separately (or even to add them). The proportion of explained variance for the multivariable model depends on the selected variables and their correlation structure. While the goodness of fit for the model can be measured by its coefficient of determination, it is far less clear what is meant by (or how to measure) the 'contribution of each variable to the outcome', let alone in percent. This is also true for the regression coefficients, as each of them depends on the selected model. In our view, the situation in NMA is similar. Indeed, NMA can be written as a meta-regression model, and so has the aforementioned issues of defining he contribution of each variable in isolation.

Random effects model

We noted that for the random effects model the omission of a study may decrease some variances, if $\hat{\tau}^2$ is not fixed. For example, for a circular network (Figure 2, left panel) leaving out a study always leads to zero inconsistency, because the resulting network is free of loops. If inconsistency was large for the primary network, the variances of all estimates decrease for the subnetworks, resulting in negative importances. In principle, we could accept negative importances: an obvious interpretation would be that a study with negative importance 'disturbs' the network. However, we prefer to fix $\hat{\tau}^2$ to the heterogeneity variance estimate from the full network, in line with the definition of the random effects weights for pairwise MA. This means that for the random effects model information of the full network enters the estimation for all subnetworks. Moreover, as $\hat{\tau}^2$ does not only depend on the variances, but also on the treatment effect estimates of all studies, information on the treatment effect estimates enters the importance values in the random effects model (as do the random effects weights in a pairwise MA). This issue is covered in detail in [5] and led Copas and others to focus mainly on common effect models [17].

Impact on the variance or on the effect estimates?

The importance of a study in a network MA can be considered from different aspects. In this article, we follow [4, 6, 7], focussing on the impact of a study on the *variances*. This type of approach generalizes the inverse variance weights in pairwise common effects MA and, like these, ignores the impact of a study on the actual *treatment effect estimates*. A study may be important because of its high precision, but this does not necessarily impact the size or direction of the effect estimates. For example, there is a marked inconsistency between the direct and the indirect estimates for comparison B:H in the thrombolytic data example: the direct effect for B:H deviates from the indirect effect, which is mainly driven by the large studies 1 and 2. Such deviations are not in the focus of variance-based methods.

For those mainly interested in treatment effects, we point to approaches to identify influential studies (or 'outliers') which impact the effect estimates [1, 2, 3]. These concepts do not rely exclusively on the structure of the network and the variances of the studies, but also account for their effect estimates and the extent to which they are consistent with estimates from other studies. We note that these methods differ from the variance-based methods in their aims.

Conclusion

We propose to measure the importance of a study for a comparison as the relative reduction of the variance of the estimate when adding the study to the network, or, equivalently, the relative loss of the precision when the study is left out. For pairwise MA, importances reduce to the usual inverse variance weights. Importances are values between 0 and 1 and cannot be expected to add up to 1. An importance of 1 means that the study is an essential link of the pathway in the network connecting one of the treatments with the other. This may possibly also hold for multiple studies on a pathway. The importance of a study for a comparison depends on the network structure and on other studies on the paths from one node in the network to another. For the common effect model, importances, like the weights in pairwise MA, are not informative with respect to the size, direction, or risk of bias of the effect estimates. This is different for the random effects model, where the effect estimates influence the importances via $\hat{\tau}^2$ like the random effect weights in pairwise MA.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
GSA had the idea to consider ratios of variances and provided a preliminary version of the R code. GR and GSC revised the R functions. GR wrote the manuscript. AN, TP, GSA, RDR, and GSC contributed to the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements
GR was funded by Deutsche Forschungsgemeinschaft (DFG), grant Nr. RU1747/1-2. GSA, AN, and TP were supported by project funding (Grant No. 179158) from the Swiss National Science Foundation.

Author details
[1]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Stefan-Meier-Strasse 26, 79104 Freiburg, Germany. [2]Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland. [3]Centre for Prognosis Research, Research Institute for Primary Care and Health Sciences, Keele University, Keele, UK.

References

1. Zhang, J., Fu, H., Carlin, B.P.: Detecting outlying trials in network meta-analysis. Statistics in Medicine **34**(19), 2695–2707 (2015)
2. Zhao, H., Hodges, J.S., Carlin, B.P.: Diagnostics for generalized linear hierarchical models in network meta-analysis. Research Synthesis Methods **8**(3), 333–342 (2017)
3. Lin, L., Chu, H., Hodges, J.S.: Alternative measures of between-study heterogeneity in meta-analysis: Reducing the impact of outlying studies. Biometrics **73**(1), 156–166 (2017)
4. Salanti, G., Giovane, C.D., Chaimani, A., Caldwell, D.M., Higgins, J.P.T.: Evaluating the quality of evidence from a network meta-analysis. PLOS ONE **9**(7), 1–14 (2014)
5. Jackson, D., White, I.R., Price, M., Copas, J., Riley, R.D.: Borrowing of strength and study weights in multivariate and network meta-analysis. Statistical Methods in Medical Research **26**(6), 2853–2868 (2017). doi: 10.1177/0962280215611702
6. Riley, R.D., Ensor, J., Jackson, D., Burke, D.L.: Deriving percentage study weights in multi-parameter meta-analysis models: with application to meta-regression, network meta-analysis and one-stage individual participant data models. Statistical Methods in Medical Research **27**(10), 2885–2905 (2018). PMID: 28162044
7. Papakonstantinou, T., Nikolakopoulou, A., Rücker, G., Chaimani, A., Schwarzer, G., Egger, M., Salanti, G.: Estimating the contribution of studies in network meta-analysis: paths, flows and streams. F1000Research **7**(610), 610 (2018). doi:10.12688/f1000research.14770.3

8. Bucher, H.C., Guyatt, G.H., Griffith, L.E., Walter, S.D.: The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of Clinical Epidemiology **50**, 683–91 (1997)
9. Dias, S., Welton, N.J., Caldwell, D.M., Ades, A.E.: Checking consistency in mixed treatment comparison meta-analysis. Statistics in Medicine **29**(7-8), 932–944 (2010)
10. König, J., Krahn, U., Binder, H.: Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Statistics in Medicine **32**(30), 5414–5429 (2013). doi: 10.1002/sim.6001
11. van Valkenhoef, G., Dias, S., Ades, A.E., Welton, N.J.: Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. Research Synthesis Methods **7**(1), 80–93 (2016)
12. White, I.R.: Network meta-analysis. Stata Journal **15**, 951–985 (2015)
13. Nikolakopoulou, A., Mavridis, D., Furukawa, T.A., Cipriani, A., Tricco, A.C., Straus, S.E., Siontis, G.C.M., Egger, M., Salanti, G.: Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: empirical study. BMJ **360** (2018)
14. Tu, Y.-K.: Node-splitting generalized linear mixed models for evaluation of inconsistency in network meta-analysis. Value in Health **19**(8), 957–963 (2016)
15. Noma, H., Tanaka, S., Matsui, S., Cipriani, A., Furukawa, T.A.: Quantifying indirect evidence in network meta-analysis. Statistics in Medicine **36**(6), 917–927 (2017)
16. Krahn, U., Binder, H., König, J.: A graphical tool for locating inconsistency in network meta-analyses. BMC Medical Research Methodology **13**, 35 (2013)
17. Copas, J.B., Jackson, D., White, I.R., Riley, R.D.: The role of secondary outcomes in multivariate meta-analysis. Journal of the Royal Statistical Society Series C-Applied Statistics **67**(5), 1177–1205 (2018). doi: 10.1111/rssc.12274
18. Jackson, D., White, I.: When should meta-analysis avoid making hidden normality assumptions? Biometrical Journal **60**(6), 1040–1058 (2018). doi:10.1002/bimj.201800071
19. Rücker, G., Krahn, U., König, J., Efthimiou, O., Schwarzer, G.: netmeta: Network Meta-Analysis using Frequentist Methods. R package version 1.1-0 (2019). https://cran.r-project.org/web/packages/netmeta/
20. White, I.R., Barrett, J.K., Jackson, D., Higgins, J.P.T.: Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Research Synthesis Methods **3**(2), 111–125 (2012)
21. Rücker, G.: Network meta-analysis, electrical networks and graph theory. Research Synthesis Methods **3**(4), 312–324 (2012)
22. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2019). R Foundation for Statistical Computing. https://www.R-project.org
23. Dias, S., Sutton, A.J., Ades, A.E., Welton, N.J.: Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical Decision Making **33**, 607–617 (2013). doi:10.1177/0272989X12458724
24. Boland, A., Dundar, Y., Bagust, A., Haycox, A., Hill, R.: Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation. Health Technology Assessment **7**(15), 1–136 (2003)
25. Lu, G.B., Ades, A.E.: Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association **101**(474), 447–459 (2006)

Figures

Figure 1 **Figure 1.** Network graph of thrombolytic data.

Figure 2 **Figure 2.** Left panel: A circle of n treatments with equal variances. Right panel: A network with a bridge.

Figure 3 **Figure 3.** Parkinson data (top left panel) with each study removed in turn (other panels).

Figure 4 **Figure 4.** A grayscale heatmap of importances for thrombolytic data. Darker colors represent greater importance of a study (column) for a comparison (row).

Tables

Additional Files

Additional file 1 — Additional File 1.R

R code to produce all analyses described in this paper.

Table 1 Importances of each study for the network with a bridge. For sake of transparency, the symbol – represents zero.

	Comparisons														
Study	A:B	A:C	B:C	C:D	D:E	D:F	E:F	A:D	B:D	C:E	C:F	A:E	A:F	B:E	B:F
AB	0.67	0.33	0.33	–	–	–	–	0.17	0.17	–	–	0.12	0.12	0.12	0.12
AC	0.33	0.67	0.33	–	–	–	–	0.44	0.17	–	–	0.36	0.36	0.12	0.12
BC	0.33	0.33	0.67	–	–	–	–	0.17	0.44	–	–	0.12	0.12	0.36	0.36
CD	–	–	–	1	–	–	–	1	1	1	1	1	1	1	1
DE	–	–	–	–	0.67	0.33	0.33	–	–	0.44	0.17	0.36	0.12	0.36	0.12
DF	–	–	–	–	0.33	0.67	0.33	–	–	0.17	0.44	0.12	0.36	0.12	0.36
EF	–	–	–	–	0.33	0.33	0.67	–	–	0.17	0.17	0.12	0.12	0.12	0.12

Table 2 Contributions of each study for the network with a bridge following the contributions approach [7]. Entries in italics differ from those in Table 1.

	Comparisons														
Study	A:B	A:C	B:C	C:D	D:E	D:F	E:F	A:D	B:D	C:E	C:F	A:E	A:F	B:E	B:F
AB	0.67	0.33	0.33	–	–	–	–	*0.11*	*0.11*	–	–	*0.07*	*0.07*	*0.07*	*0.07*
AC	0.33	0.67	0.33	–	–	–	–	*0.33*	*0.11*	–	–	*0.22*	*0.22*	*0.07*	*0.07*
BC	0.33	0.33	0.67	–	–	–	–	*0.11*	*0.33*	–	–	*0.07*	*0.07*	*0.22*	*0.22*
CD	–	–	–	1	–	–	–	*0.44*	*0.44*	*0.44*	*0.44*	*0.29*	*0.29*	*0.29*	*0.29*
DE	–	–	–	–	0.67	0.33	0.33	–	–	*0.33*	*0.11*	*0.22*	*0.07*	*0.22*	*0.07*
DF	–	–	–	–	0.33	0.67	0.33	–	–	*0.11*	*0.33*	*0.07*	*0.22*	*0.07*	*0.22*
EF	–	–	–	–	0.33	0.33	0.67	–	–	*0.11*	*0.11*	*0.07*	*0.07*	*0.07*	*0.07*

Table 3 Importance of each study for the comparisons in the Parkinson data. Direct evidence is printed in bold. The last row shows the proportion of direct evidence for each comparison in a column.

		Comparisons									
Study	Design	1:2	1:3	1:4	1:5	2:3	2:4	2:5	3:4	3:5	4:5
1	1:3	0.029	**0.531**	0.405	0.364	0.418	0.294	0.264	0.126	0.092	–
2	1:2	**0.756**	0.076	0.121	0.104	0.402	0.379	0.345	0.013	0.009	–
3	1:2:4	**0.244**	0.469	**0.595**	0.553	0.503	**0.617**	0.582	0.126	0.092	–
4	3:4	0.005	0.128	0.101	0.086	0.130	0.064	0.056	**0.535**	0.449	–
5	3:4	0.002	0.061	0.048	0.041	0.062	0.030	0.026	**0.339**	0.266	–
6	4:5	–	–	–	0.178	–	–	0.157	–	0.285	**0.577**
7	4:5	–	–	–	0.105	–	–	0.092	–	0.177	**0.423**
Direct evidence (%)		97.1	53.1	47.4	0.0	0.0	55.8	0.0	87.4	0.0	1.0

Table 4 Study weights for pairwise meta-analysis of comparisons in the Parkinson data (common effects model).

Comparison	Study	Design	Weight
1:2	2	1:2	0.779
	3	1:2:4	0.221
3:4	4	3:4	0.612
	5	3:4	0.388
4:5	6	4:5	0.577
	7	4:5	0.423

Additional file 2 — Additional File 2.pdf
Plots resulting from running all commands in Additional File 1.

Additional file 3 — Additional File 3.csv
Thrombolytic data in csv format (can be viewed as an Excel file).

Table 5 Importance of each study for the comparisons to treatment A for the thrombolytic data, multiplied by 100 for comparison to [6, Table 4 therein]. Direct evidence is printed in bold. The last row shows the proportion of direct evidence for each comparison in a column.

		Comparisons						
Study	Design	A:B	A:C	A:D	A:E	A:F	A:G	A:H
1	ABD	**82.10**	0.03	**99.50**	59.69	50.56	2.68	0.06
2	ACH	0.30	**59.39**	0.03	0.10	0.07	0.50	**96.81**
3	AC	0.00	**0.31**	0.00	0.00	0.00	0.00	0.05
4	AC	0.00	**0.13**	0.00	0.00	0.00	0.00	0.02
5	AC	0.00	**0.13**	0.00	0.00	0.00	0.00	0.02
6	AC	0.00	**38.20**	0.00	0.00	0.00	0.21	9.20
7	AC	0.00	**0.22**	0.00	0.00	0.00	0.00	0.04
8	AC	0.00	**0.38**	0.00	0.00	0.00	0.00	0.06
9	AC	0.00	**0.29**	0.00	0.00	0.00	0.00	0.05
10	AD	0.04	0.00	**0.50**	0.01	0.01	0.00	0.00
11	AF	15.78	0.00	1.57	5.70	**45.67**	0.11	0.00
12	AG	0.14	0.08	0.01	0.05	0.03	**19.35**	0.01
13	AH	0.00	0.03	0.00	0.00	0.00	0.00	**0.15**
14	AH	0.00	0.04	0.00	0.00	0.00	0.00	**0.26**
15	AH	0.00	0.03	0.00	0.00	0.00	0.00	**0.16**
16	AH	0.00	0.12	0.00	0.00	0.00	0.00	**0.75**
17	BE	0.00	0.00	0.00	100.00	0.00	0.00	0.00
18	BF	14.98	0.00	1.47	5.38	52.79	0.11	0.00
19	BF	0.11	0.00	0.01	0.03	0.67	0.00	0.00
20	BG	0.11	0.02	0.01	0.04	0.02	6.75	0.00
21	BG	0.70	0.15	0.06	0.23	0.16	31.50	0.02
22	BH	0.71	0.08	0.06	0.23	0.16	0.00	0.49
23	BH	0.45	0.05	0.04	0.15	0.10	0.00	0.31
24	CG	0.08	0.09	0.01	0.03	0.02	12.11	0.01
25	CG	0.12	0.13	0.01	0.04	0.03	17.31	0.02
26	CG	0.09	0.09	0.01	0.03	0.02	12.59	0.01
27	CH	0.00	0.13	0.00	0.00	0.00	0.00	0.32
28	CH	0.00	0.07	0.00	0.00	0.00	0.00	0.16
Direct evidence		0.82	0.99	0.98	0.00	0.46	0.19	0.89

Table 6 Study weights for the thrombolytic data (common effects model).

Comparison	Study	Design	Weight (%)
A:C	2	ACH	59.93
	3	AC	0.32
	4	AC	0.13
	5	AC	0.13
	6	AC	38.59
	7	AC	0.22
	8	AC	0.39
	9	AC	0.30
A:D	1	ABD	99.49
	10	AD	0.51
A:H	2	ACH	98.52
	13	AH	0.17
	14	AH	0.29
	15	AH	0.17
	16	AH	0.85
B:F	18	BF	68.54
	19	BF	31.46
B:G	20	BG	17.79
	21	BG	82.21
B:H	22	BH	61.26
	23	BH	38.74
C:G	24	CG	28.83
	25	CG	41.19
	26	CG	29.99
C:H	2	ACH	98.77
	27	CH	0.81
	28	CH	0.42

Figures

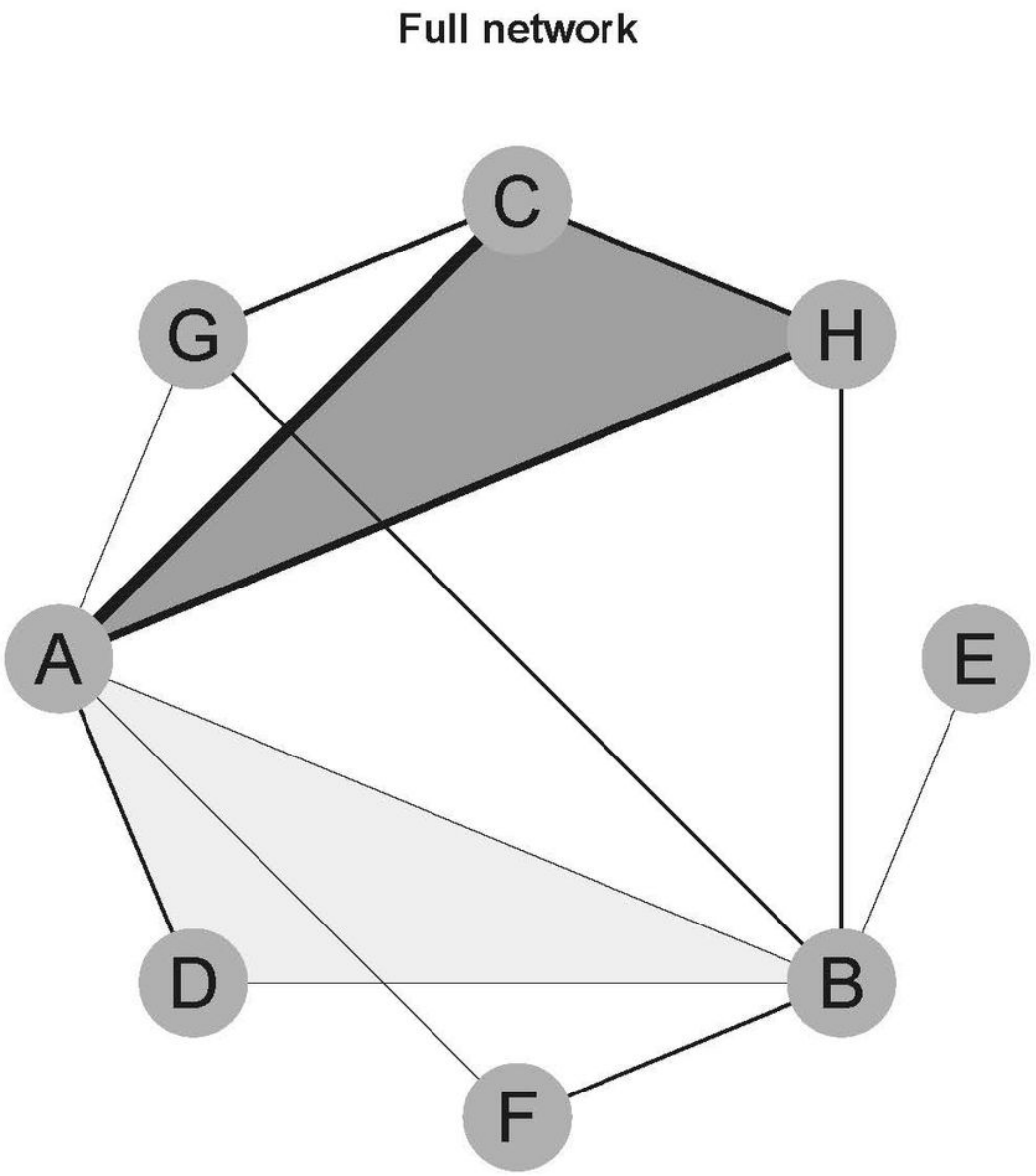

Figure 1

Network graph of thrombolytic data.

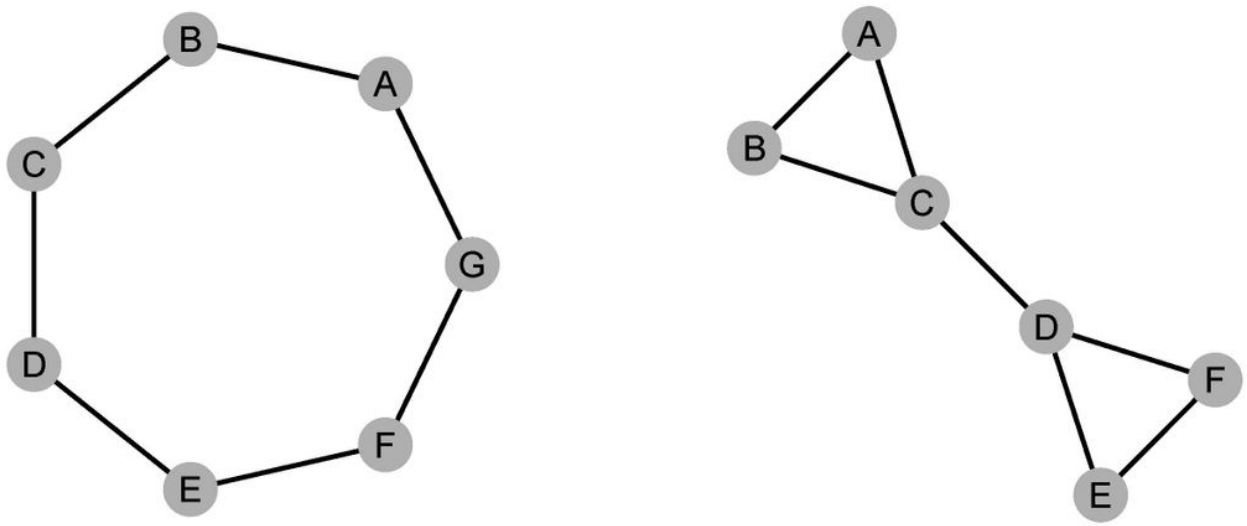

Figure 2

Left panel: A circle of n treatments with equal variances. Right panel: A network with a bridge.

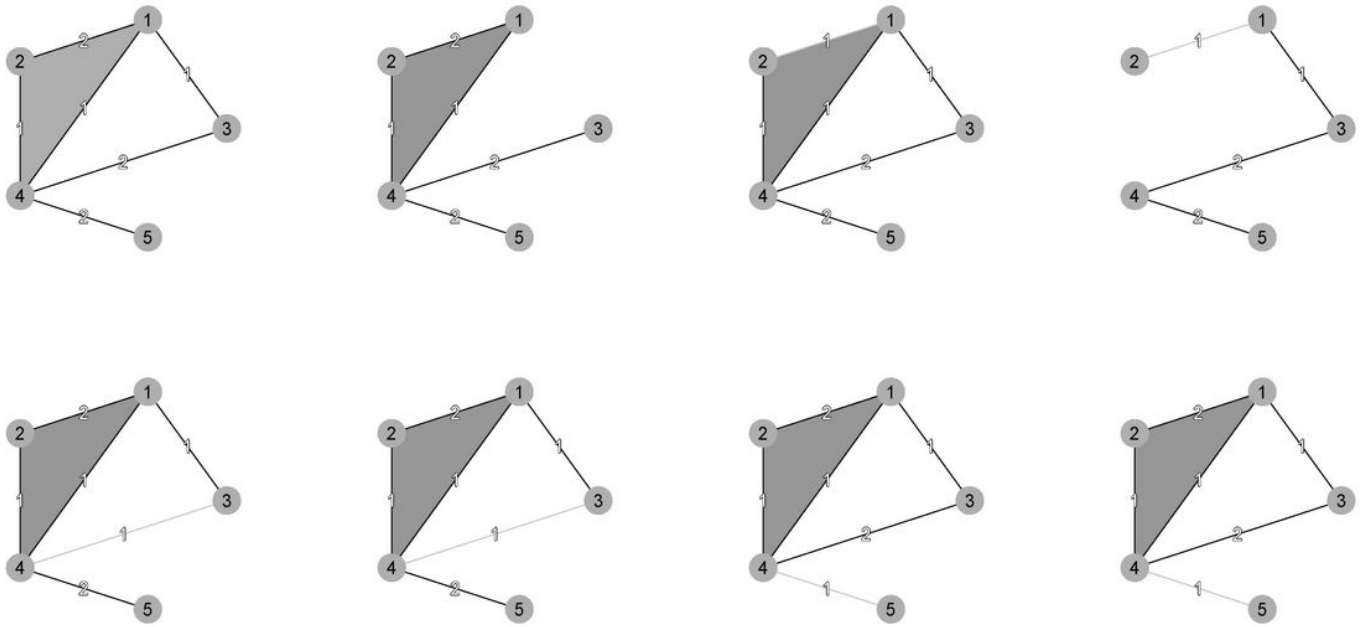

Figure 3

Parkinson data (top left panel) with each study removed in turn (other panels).

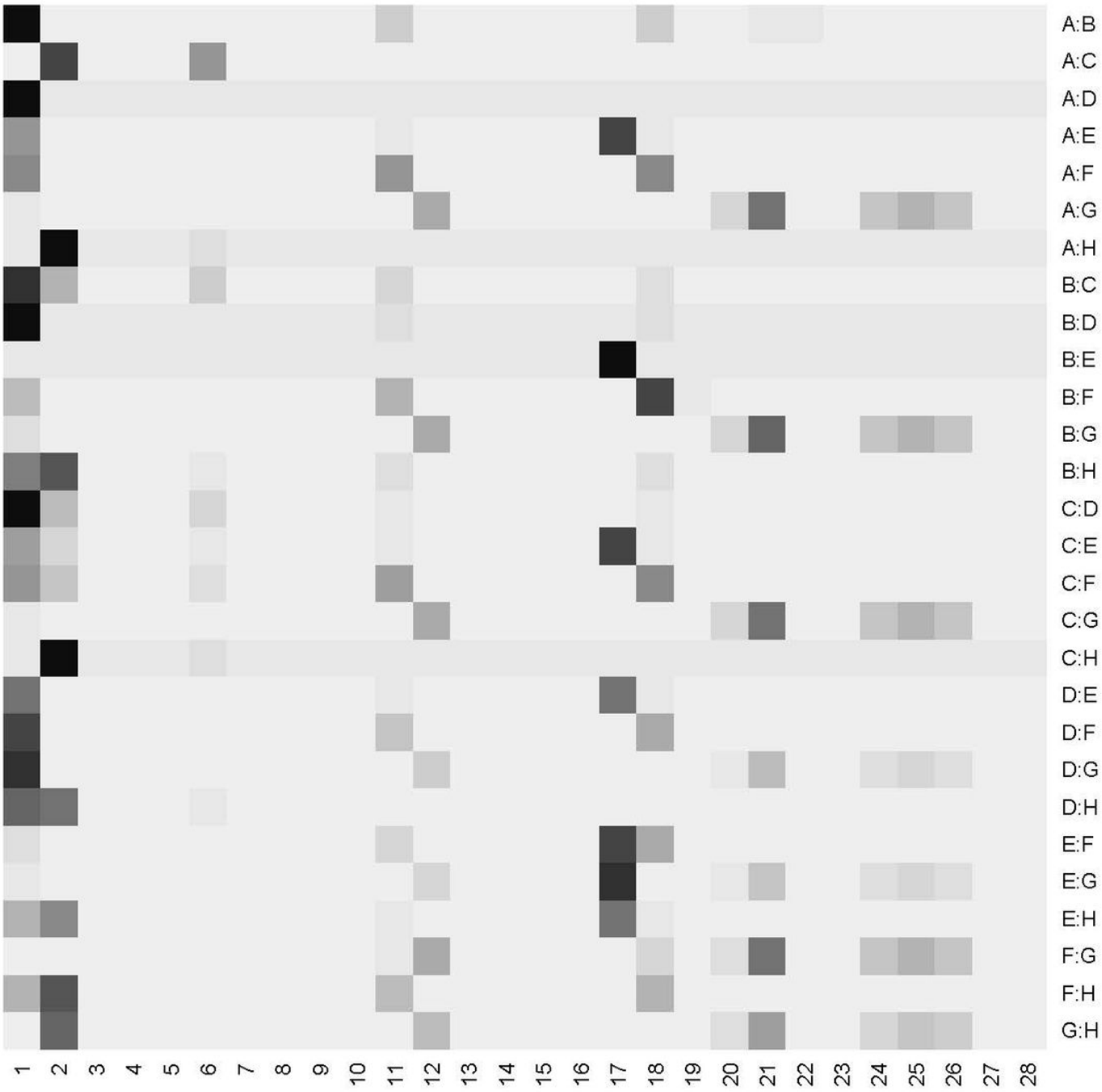

Figure 4

A grayscale heatmap of importances for thrombolytic data. Darker colors represent greater importance of a study (column) for a comparison (row).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.R
- AdditionalFile2.pdf
- AdditionalFile3.csv