Methods for including information from multi-arm trials in pairwise meta-analysis

Gerta Rücker[12], Christopher J. Cates[3] and Guido Schwarzer[1]

[1]Institute for Medical Biometry and Statistics,

Faculty of Medicine and Medical Center - University of Freiburg, Germany

[2]Cochrane Germany,

Faculty of Medicine and Medical Center - University of Freiburg, Germany

[3]Population Health Research Institute,

St George's University of London, Cranmer Terrace, London SW17 0RE, UK

July 17, 2017

Address for correspondence:

Gerta Rücker

Institute for Medical Biometry and Statistics

Stefan-Meier-Straße 26, D-79104 Freiburg, Germany

Tel. +49 (0)761 203 6673, Fax +49 (0)761 203 6680

Email: ruecker@imbi.uni-freiburg.de

Funded by Deutsche Forschungsgemeinschaft – RU1747/1-2

Conflicts of interest: None

Abstract

Keywords

Pairwise meta-analysis, multi-arm trials, multiplicity, network meta-analysis

Running head

G. Rücker et al.: Multi-arm trials in pairwise meta-analysis

Abstract: 88 words

Systematic reviewers conducting pairwise meta-analyses sometimes encounter multi-arm studies. To include these studies, in order to avoid a unit-of-analysis error, often two or more arms are combined or the control arm is split up. In this tutorial we present five different approaches that can be used. Particularly, we present a novel approach (Method 4) that to the best of our knowledge has not been presented before. We demonstrate their application on three selected data sets, discuss their scope of application and their advantages and limitations, and give recommendations.

Main text: ca. 4370 words

1 Introduction

Until ten years ago, almost all systematic reviews exclusively used standard pairwise meta-analysis. It is probably still the most frequent mode of meta-analysis in the literature. More recently, reviews have increasingly used network meta-analysis. The first Cochrane review including a network meta-analysis was published in 2010 (Walsh et al., 2010). Researchers conducting a systematic review have the choice between pairwise meta-analysis and network meta-analysis. Which method to use depends on the research question. Pairwise meta-analysis is the method of choice for researchers interested in a particular comparison, for example comparing a new treatment with a generally accepted standard of care, or asking whether a treatment is efficacious at all (compared to no treatment or placebo).

When searching the literature, reviewers planning a pairwise meta-analysis sometimes encounter multi-arm studies, where, say, two or more dosages of an interesting drug treatment are compared to a common control, for example placebo. If the research question is to compare the drug in all dosages to placebo, all comparisons (also called relative effects or contrasts) to placebo could and should be included in the same meta-analysis. However, as there is only one control group, one has to make sure that each patient enters the meta-analysis only once, in order to avoid a unit-of-analysis error. A unit-of-analysis error in a meta-analysis is said to occur if information from a treatment arm is used more than once or with disproportionate weight. For example, combining comparisons of three different dosages of an active treatment to the same full placebo group without any adjustment would mean using the patients of the placebo group thrice. This is a problem because comparisons using the same individuals are correlated. It is clearly a methodogical flaw. To avoid this, often the control group is randomly partitioned into two or more subgroups, each of which is compared to another active group. The shared group could also be the active group instead of the control group which is illustrated using a ficti-

tious example in section 16.5.4 (How to include multiple groups from one study) of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green, 2011, Version 5.1.0). The Cochrane Handbook provides general guidelines for meta-analysis with multiple treatment groups and lists a number of possible ways to include these data.

The first approach (which is recommended by Cochrane) is to combine all groups that represent different modalities of a treatment to a single large group. For the example, this results in combining two or more dosage groups. Technically, in the case of a binary outcome this is easily done by summing up the numbers of events and the total numbers of individuals across all groups involved. Section 7.7.3.8 of the Cochrane Handbook (Table 7.7.a) provides formulas for calculating mean and standard deviation of a combined group in case of a continuous outcome, given summary data for the two groups that are to be combined. This can also be done using the calculator in RevMan 5.3 (The Cochrane Collaboration, 2014).

The second approach is to split the shared group into subgroups of (nearly) equal size, one for each treatment. For the example, this is applied to the placebo group. Again, in the case of a binary outcome, this is straightforward, except that the total number of events or sample size may not be possible to split into equal parts according to the number of active treatment arms (in the extreme, we can have a study with less events in the control group than the number of active study arms). For a continuous outcome, only the total number of participants is divided up, assuming equal means and standard deviations in all subgroups. This approach is not generally recommended, as only the unit-of-analysis error is avoided, while the groups are still correlated because the means are equal.

Another approach mentioned in the Cochrane Handbook is to include two or more correlated comparisons in the meta-analysis and account for the correlation. The final proposal is to conduct a multiple treatment meta-analysis, also called network

meta-analysis, which generalizes pairwise meta-analysis to more than two treatments. In this respect, it goes beyond the scope of a systematic review that aims at comparing two interventions, based on pairwise meta-analysis. Following this recommendation of the Handbook means shifting the research question from a pairwise comparison to a multiple treatment comparison. The approach is therefore more than just another adjusting method. Because in network meta-analysis all comparisons provide information, the analysis benefits from full information, but must account for the correlation between three or more comparisons within the same trial. Appropriate methods of network meta-analysis adjust for the correlation between multiple comparisons within multi-arm studies (Franchini et al., 2012; Rücker and Schwarzer, 2014).

The objective of this tutorial is to present five approaches (three of which have been mentioned above), to demonstrate their application on three selected data sets, and to discuss their scope of application as well as their advantages and limitations. Particularly, we add a novel approach (Method 4 below) that to the best of our knowledge has not been presented before.

The paper is organised as follows. In section 2, the data examples are presented, the first with a binary outcome, the second with a continuous outcome and the third with a survival outcome. In section 3, subsection 3.1, three different potential purposes of a meta-analysis are listed. In subsection 3.2, the five methods are described and illustrated by applying them to the first example. In subsection 3.3 we discuss how to proceed if some information is missing. In section 4 all methods are applied to the other two examples. The paper concludes with a discussion and some recommendations in section 5.

2 Data sets

2.1 Pelargonium sidoides data

Our first data example comes from a Cochrane review comparing pelargonium sidoides, also known as Umckaloabo, for treatment of acute respiratory infections to placebo (Timmer et al., 2013). One of the trials included was a four-arm trial with three different treatment arms and a placebo arm (Kamin et al., 2010). The outcome is 'Failure to recover by day seven' (i.e., complete resolution of all symptoms), dichotomous data are shown in Table 1. In the Cochrane review the 'Splitting the shared group' approach was used and the trial was reported as three substudies (10 mg, 20 mg, 30 mg), each compared with a third of the number of controls. This data set serves for illustration of all methods in this section of the tutorial.

Table 1: Pelargonium sidoides data by Kamin 2010.

study	treatment	event	n
Kamin 2010	pelargonium sidoides 10 mg	91	100
Kamin 2010	pelargonium sidoides 20 mg	82	99
Kamin 2010	pelargonium sidoides 30 mg	87	99
Kamin 2010	placebo	92	101

2.2 Acupuncture data

In a three-arm trial (Linde et al., 2005), acupuncture for migraine was compared to sham acupuncture and to a waiting list. To facilitate recruitment and increase the compliance of trial physicians, a 2:1:1 randomisation ratio was used. The primary outcome was the reduction in the number of days with headache of moderate or severe intensity between the 4 weeks before and weeks 9 to 12 after randomization,

based on patients' diaries. Results are given in Table 2.

Table 2: Acupuncture data by Linde 2005

study	treatment	n	mean (days)	SD
Linde 2005	acupuncture	145	2.2	2.7
Linde 2005	sham acupuncture	81	2.2	2.7
Linde 2005	waiting list	76	0.8	2.0

2.3 Breast cancer data

The last example comes from a three-arm phase III trial investigating the efficacy and safety of three combination treatments for human epidermal growth factor receptor 2 (HER2)-negative, locally recurrent or metastatic breast cancer (Miles et al., 2010). The three arms were docetaxel 100 mg/m^2 plus bevacizumab 15 mg/kg ($n = 247$), docetaxel 100 mg/m^2 plus bevacizumab 7.5 mg/kg ($n = 248$), and docetaxel 100 mg/m^2 plus placebo ($n = 241$). The results for the primary endpoint progression-free survival (PFS) are given in Table 3.

Table 3: Breast cancer data by Miles 2010

study	treatment 1	treatment 2	HR [95% CI][1]
Miles 2010	docetaxel 100 plus bevacizumab 15	docetaxel 100 plus placebo	0.77 [0.64; 0.93]
Miles 2010	docetaxel 100 plus bevacizumab 7.5	docetaxel 100 plus placebo	0.86 [0.72; 1.04]

[1]HR = hazard ratio; CI = confidence interval

3 Methods

3.1 The purpose of the meta-analysis

When deciding how to handle multiple groups in a pairwise meta-analysis it is important to consider which is the purpose of the meta-analysis (note that the intention of the meta-analysis, as part of a systematic review, may differ from that of a primary study). The following questions are selected from an abundance of possible research questions:

1. Is treatment A efficacious, compared to placebo?
2. Is treatment A, in any application, efficacious compared to placebo?
3. Is treatment A efficacious, compared to placebo or no treatment?
4. Is A combined with B better than A alone?
5. Which of treatments A, B, C, ... is best with respect to efficacy?

Questions 1, 2 and 4 lead to pairwise meta-analysis, question 3 may suggest pairwise meta-analysis (if we are not interested in comparing placebo with no treatment), or network meta-analysis (if we are interested in comparing placebo with no treatment), only question 5 undoubtedly requires network meta-analysis.

The various types of questions may occur in different clinical areas with different probabilities. In areas with an abundance of treatments (for example, depression), network meta-analysis is the method of choice. In other fields, a research question may be much more focussed, suggesting pairwise meta-analysis, for example, when comparing a certain surgical technique with another technique, or deciding whether surgery is necessary at all (illustrated by the title of a review 'To close or not to

close: contemporary indications for patent foramen ovale closure' Zier et al. (2016)). To give these scenarios a structure, we distinguish three common cases.

1. Two treatments are compared. The distinction between different modalities of the active treatment or subtypes of the control treatment is not important for the research question at hand.

2. Two treatments are compared, but it is also of interest whether the treatment modality or the type of control treatment makes a difference.

3. The meta-analysis is designed as a network meta-analysis to determine a rank order of benefit, and the literature search was designed to find studies comparing at least two of the eligible treatments.

In the first case one would likely prefer to combine the treatment groups or control groups. This leads to our Method 1, 'Combining groups'. In the second case, there are a number of different options. First, one could split up the shared group (in example 2.1, this is the placebo group) into a number of subgroups (e.g., three placebo subgroups, corresponding to the three doses of the active treatment) and include these as substudies in the meta-analysis. We call this Method 2 ('Splitting the shared group'). In the meta-analysis context, this enables meta-regression, using the factor defining the subgroups (e.g., dosage) as a covariate. Alternatively, instead of formally splitting up the shared group, one could compare each treatment subgroup to the full group, but account for multiplicity by appropriately adjusting the standard error. There are two ways for doing this, which we call Method 3 ('Approximate adjustment') and Method 4 ('Exact adjustment'). In the third common case, it is possible to conduct a network meta-analysis, which we call Method 5 ('Network meta-analysis'). In the next subsection we describe all five methods in more detail and illustrate each of them practically by application to the first data set, 2.1.

3.2 Methods for including multiple groups into a meta-analysis

Method 1: 'Combining groups'

For a binary outcome, as in the first example data set, combining, i.e., merging the groups simply means adding the numbers of events and total participants over all groups. This provides the data in Table 4 and leads to a risk ratio (RR) of 0.96 with a 95% confidence interval (CI) of [0.89; 1.03] (inverse variance method).

Table 4: 'Combining groups' applied to the pelargonium sidoides data.

	active treatment		control	
	pelargonium sidoides (10/20/30 mg)		placebo	
study	events	n	events	n
Kamin 2010	260	298	92	101

Section 7.7.3.8 of the Cochrane Handbook (Table 7.7.a) provides formulas for calculating mean and standard deviation in case of a continuous outcome, if summary data over all groups or individual patient data are not available. Given the sample sizes n_1, n_2, the means $\bar{x}_1, \bar{x}_2$ and the standard deviations $\mathrm{SD}_1, \mathrm{SD}_2$, we obtain for the combined group

$$n_{comb} = n_1 + n_2 \tag{1}$$

$$\bar{x}_{comb} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \tag{2}$$

$$\mathrm{SD}_{comb} = \sqrt{\frac{(n_1-1)\mathrm{SD}_1^2 + (n_2-1)\mathrm{SD}_2^2 + \frac{n_1 n_2}{n_1+n_2}(\bar{x}_1 - \bar{x}_2)^2}{n_1 + n_2 - 1}}. \tag{3}$$

If we have arm-based survival data for two (or more) arms in form of Kaplan-Meier curve estimates or numbers of events and numbers at risk for all time points, these

arms can be combined in a straightforward way. For detailed methods of extraction for survival data, see Parmar et al. (1998); Tierney et al. (2007).

Method 2: 'Splitting the shared group'

This method was applied in the Cochrane review that included the study of the first example. The control group was randomly split in three groups, leading to the data in Table 5. Meta-analysis using the common effect model with the Mantel-Haenszel method resulted in a pooled RR of 0.96 with a 95% CI of [0.89; 1.03], which is in agreement with the result of method 1. Using the inverse variance method in analogy to Method 1 leads to a pooled RR of 0.96 [0.89; 1.04]. Due to the different weighting, there is a slight difference seen in the confidence interval.

Table 5: 'Splitting the shared group' applied to the pelargonium sidoides data.

	active treatment		control	
	pelargonium sidoides		placebo	
study	events	n	events	n
Kamin 2010, 10 mg	91	100	31	34
Kamin 2010, 20 mg	82	99	30	33
Kamin 2010, 30 mg	87	99	31	34

Method 3: 'Approximate adjustment'

This method has been proposed by one of the authors of this tutorial (Cates, 2015), and can be used to avoid unit of analysis errors when including multi-arm trials in a generic inverse variance meta-analysis. It is based on the following consideration. Instead of formally splitting up the shared group into groups of (nearly) equal size and (nearly) equal numbers of events, we could compare each active group to the full

control group, but increase the standard error of this comparison by an appropriate factor. This can be derived as follows.

Let k be the total number of arms in the study including the control arm. In our example, we have $k = 4$. Comparing each active arm to the kth group (the control group), we observe variance estimates v_{ik} for all $k-1$ pairwise comparisons of arm i to arm k. These can be assumed to be sums of the respective arm-based sampling variances s_i^2 and s_k^2, that is

$$v_{ik} = s_i^2 + s_k^2 \ (i = 1, \ldots, k-1).$$

Splitting the control group into $k-1$ equal parts means increasing the variance in the control group, s_k^2, by the factor $k-1$. Accounting for multiple use of the control group therefore means that we have to consider the adjusted variances

$$v_{ik}^{adj} = s_i^2 + (k-1)s_k^2 \ (i = 1, \ldots, k-1)$$

instead of v_{ik}. The variance of the naïvely pooled common effect estimate is

$$\text{Var}^{\,unadj} = \frac{1}{\sum_{i=1}^{k-1} \frac{1}{v_{ik}}} = \frac{1}{\sum_{i=1}^{k-1} \frac{1}{s_i^2 + s_k^2}}$$

and the variance of the pooled common effect estimate based on the adjusted variances is

$$\text{Var}^{\,adj} = \frac{1}{\sum_{i=1}^{k-1} \frac{1}{v_{ik}^{adj}}} = \frac{1}{\sum_{i=1}^{k-1} \frac{1}{s_i^2 + (k-1)s_k^2}}.$$

We obtain for the variance inflation factor

$$\frac{\text{Var}^{\,adj}}{\text{Var}^{\,unadj}} = \frac{\sum_{i=1}^{k-1} \frac{1}{s_i^2 + s_k^2}}{\sum_{i=1}^{k-1} \frac{1}{s_i^2 + (k-1)s_k^2}}.$$

If all arms including the control arm are of approximately equal size, we may estimate the variance inflation factor by assuming that all within-group standard errors are equal, that is, $s_i^2 = s^2 (i = 1, \ldots, k)$. Then the approximate variance inflation factor becomes

$$\frac{\text{Var}^{\,adj}}{\text{Var}^{\,unadj}} = \frac{k}{2}$$

and the inflation factor for the standard error is $\sqrt{k/2}$. Applying this to the example ($k = 4$) means first comparing all active arms to the (unsplit) control arm by calculating pairwise RRs and then inflating the standard error of each RR by $\sqrt{2}$ before performing meta-analysis. For the example, this results in a pooled RR of 0.96 with a 95% CI of [0.89; 1.04], which is in accordance to those of methods 1 and 2 (Figure 1).

We note that Method 3 can be applied to a more general situation. Let us consider a k-arm study, where k_T arms are interpreted as active arms and k_C arms as control arms ($k = k_T + k_C$). The aim is to compare 'active' to 'control' in a pairwise meta-analysis, given all $k_T \times k_C$ pairwise comparisons of active treatments with control treatments, like in Figure 2 with $k_T = 2,\ k_C = 3$. Of course, simply pooling all comparisons in a meta-analysis would lead to a unit-of-analysis error. Assuming equal standard errors in all arms, we have for each of the $k_T \times k_C$ contrasts a sampling variance of $s^2 + s^2 = 2s^2$ (without adjustment).

Splitting each arm using Method 2 ('Splitting the shared group') would lead to dividing each of the active arms into k_C equal parts (one for each control arm) and each of the control arms into k_T equal parts (one for each active arm). Adjusting the variances thus leads to multiplying the variance of each active arm by k_C and multiplying the variance of each control arm by k_T. Accordingly, each comparison obtains the variance $k_C s^2 + k_T s^2 = k s^2$. Compared to naive pooling without adjustment, we again have a variance inflation factor of $ks^2/(2s^2) = k/2$ for the variance und $\sqrt{k/2}$ for the standard error.

Method 4: 'Exact adjustment'

Method 3 is an approximate method, because it is based on the simplifying assumption that the standard errors in all arms were equal. However, this might not be the

case if the arms are of different size. It is particularly questionable in the case of binary data if the number of events varies across arms, as the variance also depends on the number of events. To overcome this, we propose to adjust the standard errors by exact inflation factors that can be derived using a method that was developed for adjusting the standard errors of multi-arm studies in network meta-analysis (Rücker and Schwarzer, 2014). The method basically reverses what happens with the variances in a multi-arm study. Whereas precision in multi-arm studies benefits from multiple comparisons, we have to reverse this when using only a part of the comparisons.

For a k-arm study we have $k(k-1)/2$ comparisons. We consider the variances of all pairwise comparisons of arms i and j with arm-based sampling variances s_i^2 and s_j^2,

$$v_{ij} = s_i^2 + s_j^2.$$

For the k-arm study we now define the full design matrix $\mathbf{X}$ with dimension $k(k-1)/2 \times k$. As all effects are determined by the $k-1$ contrasts to an arbitrary reference treatment, $\mathbf{X}$ has rank $k-1$. For example, for $k = 4$ we have $k(k-1)/2 = 6$ and

$$\mathbf{X} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \tag{4}$$

with rank 3. Let $\mathbf{V}$ be a $k \times k$ matrix containing the $k(k-1)/2$ contrast-based

sampling variances v_{ij}, ordered in the following way (example for $k = 4$):

$$\mathbf{V} = \begin{pmatrix} 0 & v_{12} & v_{13} & v_{14} \\ v_{12} & 0 & v_{23} & v_{24} \\ v_{13} & v_{23} & 0 & v_{34} \\ v_{14} & v_{24} & v_{34} & 0 \end{pmatrix}. \tag{5}$$

Then the matrix $\mathbf{L}$ defined by

$$\mathbf{L} = \left(-\frac{1}{2k^2} \mathbf{X}^\top \mathbf{X} \mathbf{V} \mathbf{X}^\top \mathbf{X} \right)^+$$

(where the + symbol denotes the pseudoinverse operator (Albert, 1972)) contains as off-diagonal elements the negative inverses of the inflated variances, ordered in the same way as in $\mathbf{V}$ (Xiao and Gutman, 2004; Rücker, 2012; Rücker and Schwarzer, 2014).

For arbitrary k we provide R code, based on the R package netmeta, in Appendix A of this article (R Core Team, 2014; Rücker et al., 2017). For $k = 3$, the inflated variances v_{ij}^{adj} can also be calculated by hand, see Appendix B. Applying exact adjustment to the first example leads to a pooled RR of 0.97 with a 95% CI of [0.90; 1.04] (Figure 3).

We note that Method 4 not only corrects the standard errors, but also adjusts for the correlation between multiple comparisons within the same study (Rücker and Schwarzer, 2014).

Method 5: 'Network meta-analysis'

Network meta-analysis is an extension of pairwise meta-analysis, used to compare three or more treatments for a given medical condition, based on combining information from all existing comparisons among the treatments from a number of studies (Salanti, 2012; Rücker, 2016). Typically, also multi-arm studies are included. One

or more studies comparing different dosages (say) to each other and/or placebo can be pooled in a network meta-analysis, providing amalgamated direct and indirect evidence for each possible comparison. For the pelargonium sidoides data, network meta-analysis would be of particular interest if there were more studies reporting results stratified by dosage. Figure 4 shows the result for our pelargonium sidoides study, interpreted as a network meta-analysis of a single study. The standard errors are not adjusted, as the arms are not merged or split. We do not go into more detail here, as this is beyond the scope of this article in which our main focus is on pairwise meta-analysis.

3.3 Missing information

In the pelargonium sidoides example we have full information for all arms. In practice, however, information may be missing, particularly in case of a continuous outcome. Often only contrast-based variances are known, and not all of them are reported in the primary studies. In this subsection, we consider such situations and show which methods can be chosen.

All contrast-based sampling variances known

If variances are provided for all possible contrasts (comparisons), we know matrix $\mathbf{V}$ as in equation (5) and can use methods 3 (approximate adjustment) and 4 (exact adjustment). In this case it is possible to derive all arm-based variances, as shown in Appendix C. If the treatment effects are also available for all contrasts, one can set one arm effect, considered as a baseline (for example, the placebo effect), formally to zero (or any other value, because this value cancels out). Then all other arm-based responses are determined, and we can continue as if having full information and also use methods 1 (combining groups), 2 (splitting the shared group), and 5 (network

meta-analysis).

Only a subset of the contrast-based sampling variances known

If not all contrast-based sampling variances are known, but only those that compare each treatment to the control, we either may use method 3 (approximate adjustment) or use method 4 (exact adjustment) after imputing missing information. We show how the latter works for $k = 3$. Assume we know the variances v_{13} and v_{23}. Then we may estimate v_{12} by $v_{12} = (v_{13} + v_{23})/2$ and use exact adjustment.

Alternatively, and particularly if the arms are of different sizes n_1, n_2, n_3, we could suppose a common standard deviation SD for all arms and assume

$$v_{12} = \frac{\mathrm{SD}^2}{n_1} + \frac{\mathrm{SD}^2}{n_2} \tag{6}$$

$$v_{13} = \frac{\mathrm{SD}^2}{n_1} + \frac{\mathrm{SD}^2}{n_3} \tag{7}$$

$$v_{23} = \frac{\mathrm{SD}^2}{n_2} + \frac{\mathrm{SD}^2}{n_3} \tag{8}$$

Equations 7 and 8 provide

$$\mathrm{SD}^2 = v_{13}\frac{n_1 n_3}{n_1 + n_3}$$
$$\mathrm{SD}^2 = v_{23}\frac{n_2 n_3}{n_2 + n_3}$$

and we may estimate SD^2 as a mean from these:

$$\mathrm{SD}^2 = \frac{1}{2}\left(v_{13}\frac{n_1 n_3}{n_1 + n_3} + v_{23}\frac{n_2 n_3}{n_2 + n_3}\right)$$

Inserting this in (6) we obtain

$$v_{12} = \frac{n_3}{2}\left(\frac{n_1 + n_2}{n_1 n_2}\right)\left(\frac{n_1 v_{13}}{n_1 + n_3} + \frac{n_2 v_{23}}{n_2 + n_3}\right). \tag{9}$$

Table 6 summarises when each method can be applied.

Table 6: Required data for application of the adjustment methods.

	Available data		
Method	Arm-based response data for all arms	Contrast-based data for all contrasts	Contrast-based data only for the contrast to the shared group
Combining groups	yes	yes[1]	no
Splitting the shared group	yes	yes[1]	no
Approximate adjustment	yes	yes	yes
Exact adjustment	yes	yes	yes, imputation necessary
Network meta-analysis	yes	yes	yes, imputation necessary

[1]Setting one arm effect, considered as a baseline, formally to zero

4 Application to the other data sets

In this section we apply the different methods to the other two example data sets.

4.1 Acupuncture data

Suppose that a meta-analysis of 'all types of acupuncture versus no treatment' would consider studies of either 'acupuncture versus waiting list' or studies of 'sham acupuncture versus waiting list' to be eligible for inclusion.

Method 1: 'Combining groups'

As the outcome 'reduction in days with moderate or severe headache' is a continuous outcome, we have to use the formulas given in section 7.7.3.8 of the Cochrane Handbook (Table 7.7.a) for combining the acupuncture and the sham acupuncture group. Given the sample sizes n_1, n_2, the means $\bar{x}_1, \bar{x}_2$ and the standard deviations

$\mathrm{SD}_1, \mathrm{SD}_2$ from Table 2 and using equations (1) to (3), we obtain for the combined group $n_{comb} = 226, \bar{x}_{comb} = 2.2, \mathrm{SD}_{comb} = 2.69$. Comparing this to the waiting list group provides a mean difference of 1.4 with a 95% CI of [0.83; 1.97].

Method 2: 'Splitting the shared group'

Splitting the waiting list group leads to two formal waiting list groups each having 38 patients with means 0.8 and standard deviations 2. Comparing the acupuncture group to one of them and the sham acupuncture group to the other and pooling this in a meta-analysis results in a mean difference of 1.4 with a 95% CI of [0.82; 1.98] (Figure 5).

Method 3: 'Approximate adjustment'

Approximate adjustment means multiplying the standard errors for each of the comparisons, acupuncture vs waiting list and sham acupuncture vs waiting list, with the factor $\sqrt{3/2}$ and pooling the results in a meta-analysis. This results in a mean difference of 1.4 with a 95% CI of [0.81; 1.99] (Figure 6).

Method 4: 'Exact adjustment'

Choosing method 4 'Exact adjustment', we can use the R package netmeta to calculate the exact inflated standard errors (R code see Appendix A), or, alternatively, we can calculate the inflated standard errors using the formulas in Appendix B. Using the latter method we obtain

$$S = 0.5(2 \cdot 0.37^2 \cdot 0.32^2 + 2 \cdot 0.37^2 \cdot 0.38^2 + 2 \cdot 0.32^2 \cdot 0.38^2 - 0.37^4 - 0.32^4 - 0.38^4) = 0.02$$

and the variances

$$
\begin{aligned}
v_{12}^{adj} &= \frac{S}{-0.37^2 + 0.32^2 + 0.38^2} && = 0.23 \\
v_{13}^{adj} &= \frac{S}{0.37^2 - 0.32^2 + 0.38^2} && = 0.13 \\
v_{23}^{adj} &= \frac{S}{0.37^2 + 0.32^2 - 0.38^2} && = 0.24.
\end{aligned}
$$

This leads to adjusted standard errors $\sqrt{v_{13}^{adj}} = 0.36$ and $\sqrt{v_{13}^{adj}} = 0.49$ for the comparisons 'acupuncture versus waiting list' and 'sham acupuncture versus waiting list'. Combining both groups with the adjusted standard errors in a meta-analysis results in a mean difference of 1.4 with a 95% CI of [0.83; 1.97] (Figure 7). We note that the exact variance inflation factors for the three comparisons (1.61, 1.29, 1.66) differ from their approximate value of 3/2 due to the unbalanced group sizes. For this reason the exact adjustment method best reproduces the result from the 'Combining groups' approach.

Method 5: 'Network meta-analysis'

Figure 8 shows the results of network meta-analysis applied to the acupuncture data, with 'waiting list' as reference. As for the pelargonium sidoides data, this analysis simply reproduces the result of the three-arm study without adjustment, as no merging or splitting of groups is necessary or intended.

4.2 Breast cancer data

The breast cancer data are an example of missing information. We have only contrast-based information (hazard ratios) for two of the three comparisons, docetaxel 100 + bevacizumab 15 and 7.5, respectively, vs docetaxel 100 + placebo. Thus, to compare bevacizumab in either dosage to docetaxel alone, the arm-based methods 1 and 2 are not possible. We may use the contrast-based methods 3 (ap-

proximate adjustment) or 4 (exact adjustment, that requires imputation of further information, using methods from subsection 3.3).

Method 3: 'Approximate adjustment'

For the given contrasts, we can use the estimated hazard ratios and the confidence intervals to derive the log hazard ratios $\ln 0.77 = -0.2614$ and $\ln 0.86 = -0.1508$ with standard errors 0.0953 and 0.0938. Multiplying the standard errors with $\sqrt{3/2}$ provides adjusted standard errors 0.1168 and 0.1149. Using these in a meta-analysis to pool the bevacizumab arms results in a pooled HR of 0.81 with 95% CI of [0.69; 0.96] (Figure 9).

Method 4: 'Exact adjustment'

The variance of the missing comparison (bevacizumab 15 vs bevacizumab 7.5) can be estimated using equation (9). We obtain 0.0939 for the missing contrast-based standard error. We may now use the R code in appendix A or the formulas in appendix B for exact adjustment of the standard errors, giving 0.1180 for bevacizumab 15 vs docetaxel alone and 0.1142 for bevacizumab 7.5 vs docetaxel alone. Using these in a meta-analysis to pool the bevacizumab arms results in a pooled HR of 0.82 with 95% CI of [0.69; 0.96] (Figure 10).

Method 5: 'Network meta-analysis'

Figure 11 shows the results of network meta-analysis applied to the breast cancer data, with docetaxel 100 + placebo as reference. Again, this analysis almost reproduces the result of the three-arm study.

5 Discussion

Pairwise meta-analysis still dominates the literature of systematic reviews. Even in reviews using network meta-analysis as primary analysis, pairwise meta-analyses are routinely conducted and also additionally requested by recent guidelines (Puhan et al., 2014; Caldwell et al., 2016). In addition to network meta-analysis, which is designed for comparing more than two treatments in a systematic review, we have presented four methods for including multiple arms from a multi-arm study in a pairwise meta-analysis. Some type of adjustment is necessary when doing this to avoid a unit-of-analysis error.

We distinguished different purposes for including multiple arms. If some arms are to be combined, as their distinction is not relevant for the research question at hand, the 'Combining groups' method can be chosen. If one decides to analyse the arms separately, one should choose one of the other methods (splitting the shared group, approximate adjustment, exact adjustment, or network meta-analysis). Whereas approximate adjustment is always possible, group-splitting, exact adjustment and network meta-analysis make some requirements with respect to data availability. We showed how missing information can be imputed, if necessary.

If the common effect model (traditionally called fixed effect model) is used for pooling, the relative weight of a multi-arm study does not depend on whether its arms are combined (Method 1) or split up (Methods 2, 3, and 4). When using the random effects model, however, the relative weight of each study may depend on the method for adjusting. Suppose, for example, that there is heterogeneity between the comparisons of different active arms of a multi-arm study to a common control. This heterogeneity is hidden if these arms are combined (Method 1), but it becomes apparent if the control group is split up (Methods 2, 3, and 4). Accordingly, the estimate for the between-study variance τ^2 may change when using one of these

methods, which in turn affects all random effect weights, also those of other studies. Furthermore, if a control arm is split up into two or more parts, τ^2 enters more than one line in the data, and the study will effectively obtain more weight in the meta-analysis (particularly if τ^2 is large compared to the study variance) in comparison to combining the active arms. For this reason, we explicitly discourage using Methods 2, 3, and 4 when using the random effects model.

Conclusion

If a pairwise meta-analysis is planned based on the common effect model, we recommend using the 'Combining groups' or the 'Exact adjustment' approach, whenever possible. The 'Splitting the shared group' approach seems somewhat arbitrary. In our examples the 'Approximate adjustment' method also provided similar results to the 'Combining groups' and the 'Exact adjustment' approaches. Methods 'Splitting the shared group', 'Approximate adjustment' and 'Exact adjustment' should be avoided when using the random effects model. When planning a systematic review where several treatments are available, or treatments are available in different dosages, pharmaceutical forms or settings, network meta-analysis should always be considered as an alternative to pairwise meta-analysis.

A R code for approximate and exact adjustment

In practice, if all variances are given or imputed, determining inflation factors for arbitrary k is possible using the R package `netmeta` for network meta-analysis. We show R code for the pelargonium sidoides example.

```
# Install package netmeta and make it available in R session
install.packages("netmeta")
library(netmeta)
# Data set by Kamin et al. (2010)
kamin <- data.frame(study = rep("Kamin 2010", 4),
                    treatment = c("10 mg", "20 mg", "30 mg", "placebo"),
                    n.failures = c(91, 82, 87, 92),
                    n.patients = c(100, 99, 99, 101))
# Use pairwise() to obtain pairwise contrasts
# with risk ratio (RR) as effect measure
p.kamin <- pairwise(event = n.failures, n = n.patients,
                    treat = treatment, studlab = study,
                    data = kamin,
                    sm = "RR")
p.kamin
# Conduct network meta-analysis
# (to calculate adjusted standard errors)
net <- netmeta(TE, seTE, treat1, treat2, studlab,
               data = p.kamin)
# Print adjusted standard errors:
as.data.frame(net)[, 1:6]
# Forest plot (Figure 4)
forest(net, ref = "placebo", leftlab = "Comparison to placebo")
# Calculate approximately adjusted standard errors
```

```r
# with inflation factor sqrt(n / 2) = sqrt(2)
infl <- sqrt(net$n / 2)
seTE.approx <- infl * net$seTE
# Print unadjusted and adjusted standard errors
data.frame(comparison  = rownames(net$Cov.fixed),
           unadjusted  = net$seTE,
           approximate = seTE.approx,
           exact       = net$seTE.adj)
```

B Determining inflated variances for $k = 3$

In this subsection, we show how to obtain exactly adjusted standard errors by hand for $k = 3$. We have

$$\mathbf{L} = \frac{1}{S}\begin{pmatrix} 2v_{23} & v_{12} - v_{13} - v_{23} & -v_{12} + v_{13} - v_{23} \\ v_{12} - v_{13} - v_{23} & 2v_{13} & -v_{12} - v_{13} + v_{23} \\ -v_{12} + v_{13} - v_{23} & -v_{12} - v_{13} + v_{23} & 2v_{12} \end{pmatrix}$$

with

$$S = \frac{1}{2}\left(2v_{12}v_{13} + 2v_{12}v_{23} + 2v_{13}v_{23} - v_{12}^2 - v_{13}^2 - v_{23}^2\right).$$

The inflated variances v_{ij}^* are obtained from the negative inverses of the off-diagonal:

$$v_{12}^* = \frac{S}{-v_{12} + v_{13} + v_{23}}$$
$$v_{13}^* = \frac{S}{v_{12} - v_{13} + v_{23}}$$
$$v_{23}^* = \frac{S}{v_{12} + v_{13} - v_{23}}$$

The adjusted standard errors are obtained by taking the square roots of $v_{12}^*, v_{13}^*, v_{23}^*$. The weight reduction factors f_{12}, f_{13}, f_{23} are given by

$$f_{12} = \frac{v_{12}}{v_{12}^*} = \frac{v_{12}(-v_{12} + v_{13} + v_{23})}{S}$$
$$f_{13} = \frac{v_{13}}{v_{13}^*} = \frac{v_{13}(v_{12} - v_{13} + v_{23})}{S}$$
$$f_{23} = \frac{v_{23}}{v_{23}^*} = \frac{v_{23}(v_{12} + v_{13} - v_{23})}{S}.$$

The average weight reduction factor is

$$\frac{1}{3}(f_{12}+f_{13}+f_{23}) = \frac{-v_{12}^2 - v_{13}^2 - v_{23}^2 + 2v_{12}v_{13} + 2v_{12}v_{23} + 2v_{13}v_{23}}{3S} = \frac{2S}{3S} = \frac{2}{3},$$

not depending on the variances. This again leads to the approximate adjusting method and is a special case of the general case of k arms, where the average weight reduction factor is $2/k$ (Rücker 2012). The inverse of the average weight reduction factor, here 3/2 (in general $k/2$) is thus the harmonic mean of the variance inflation factors.

C Determining arm-based variances from contrast-based variances

If variances are provided for all possible contrasts, we can derive arm-based variances from these. This works as follows. For a k-arm study, the design matrix $\mathbf{X}$ is that of a complete graph of k vertices and $k(k-1)/2$ edges as in equation (4). Taking the first k rows of its absolute version $|\mathbf{X}|$, we obtain a quadratic matrix, such as

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

with rank k (here $k = 4$). Taking the first k of the contrast-based variances, arranged in the same order as in matrix $\mathbf{V}$ (equation(5)), i.e., $\mathbf{v} = (v_{12}, v_{13}, v_{14}, v_{23})$, we can derive the underlying arm-based variances $\mathbf{s} = (s_1^2, \ldots, s_k^2)^\top$ by solving the equation $\mathbf{As} = \mathbf{v}$ which yields

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{v}.$$

References

Albert, A. E., 1972. Regression and the Moore-Penrose Pseudoinverse. Mathematics in Science and Engineering. Academic Press, New York, ISBN: 0-12-048450-1.

Caldwell, D. M., Ades, A. E., Dias, S., Watkins, S., Li, T., Taske, N., Naidoo, B., Welton, N. J., Jul 16 2016. A threshold analysis assessed the credibility of conclusions from network meta-analysis. Journal of Clinical Epidemiology 80, 68–76, doi: 10.1016/j.jclinepi.2016.07.003.

Cates, C., 2015. Multiple-arm trial data: using a corrected standard error for GIV analyses. Cochrane Database of Systematic Reviews Suppl 1-327.
URL http://www.cochranelibrary.com/dotAsset/5bc8d6fd-0604-4bed-beb2-8bc7a74aa4b9.pdf

Franchini, A. J., Dias, S., Ades, A. E., Jansen, J. P., Welton, N. J., July 20 2012. Accounting for correlation in network meta-analysis with multi-arm trials. Research Synthesis Methods 3 (2), 142–160, dOI: 10.1002/jrsm.1049.

Higgins, J. P., Green, S., 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration.
URL http://www.cochrane-handbook.org

Kamin, W., Maydannik, V. G., Malek, F. A., Kieser, M., 2010. Efficacy and tolerability of EPs 7630 in patients (aged 6-18 years old) with acute bronchitis. Acta Paediatrica 99 (4), 537–543.

Linde, K., Streng, A., Jürgens, S., Hoppe, A., Brinkhaus, B., Witt, C., Wagenpfeil, S., Pfaffenrath, V., Hammes, M., Weidenhammer, W., Willich, S., Melchart, D., 2005. Acupuncture for patients with migraine: a randomized controlled trial. JAMA 293 (17), 2118–2125.

Miles, D. W., Chan, A., Dirix, L. Y., Cortés, J., Pivot, X., Tomczak, P., Delozier, T., Sohn, J. H., Provencher, L., Puglisi, F., Harbeck, N., Steger, G. G., Schneeweiss, A., Wardley, A. M., Chlistalla, A., Romieu, G., Jul 10 2010. Phase III study of bevacizumab plus

docetaxel compared with placebo plus docetaxel for the first-line treatment of human epidermal growth factor receptor 2-negative metastatic breast cancer. Journal of Clinical Oncology 28 (20), 3239–3247, doi: 10.1200/JCO.2008.21.6457.

Parmar, M. K. B., Torri, V., Stewart, L., 1998. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Statistics in Medicine 17, 2815–2834.

Puhan, M. A., Schünemann, H. J., Murad, M. H., Li, T., Brignardello-Petersen, R., Singh, J. A., Kessels, A. G., Guyatt, G. H., GRADE Working Group, Sep 24 2014. A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. British Medical Journal 349, g5630, doi: 10.1136/bmj.g5630.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.R-project.org

Rücker, G., 2012. Network meta-analysis, electrical networks and graph theory. Research Synthesis Methods 3 (4), 312–324.

Rücker, G., 2016. Network meta-analysis. Wiley StatsRef: Statistics Reference Online, pp. 1–8, stat07909.
URL http://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat07909/abstract

Rücker, G., Schwarzer, G., 2014. Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. Statistics in Medicine 33, 4353–4369, dOI: 10.1002/sim.6236.

Rücker, G., Schwarzer, G., Krahn, U., König, J., 2017. netmeta: Network meta-analysis using frequentist methods. R package version 0.9-3.
URL http://cran.at.r-project.org/web/packages/netmeta/

Salanti, G., June 11 2012. 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 3 (2), 80–97. URL doi:10.1002/jrsm.1037

The Cochrane Collaboration, 2014. Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre.

Tierney, J. F., Stewart, L. A., Ghersi, D., Burdett, S., Sydes, M. R., Jun 7 2007. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 8, 16.

Timmer, A., Günther, J., Motschall, E., Rücker, G., Antes, G., Kern, W. V., 2013. Pelargonium sidoides extract for treating acute respiratory tract infections. Cochrane Database of Systematic Reviews (10).

Walsh, T., Worthington, H. V., Glenny, A.-M., Appelbe, P., Marinho, V. C., Shi, X., Jan 20 2010. Fluoride toothpastes of different concentrations for preventing dental caries in children and adolescents. Cochrane Database of Systematic Reviews Doi: 10.1002/14651858.CD007868.pub2.

Xiao, W. J., Gutman, I., 2004. Relations between resistance and Laplacian matrices and their applications. Match-Communications in Mathematical and in Computer Chemistry (51), 119–127.

Zier, L., Sievert, H., Mahadevan, V., Nov. 2016. To close or not to close: contemporary indications for patent foramen ovale closure. Expert Rev Cardiovasc Ther. 14 (11), 1235–1244, doi: 10.1080/14779072.2016.1224178.

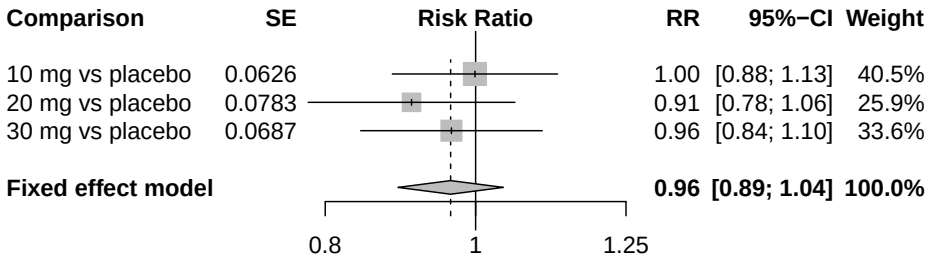

Figure 1: 'Approximate adjustment' for the pelargonium sidoides data.

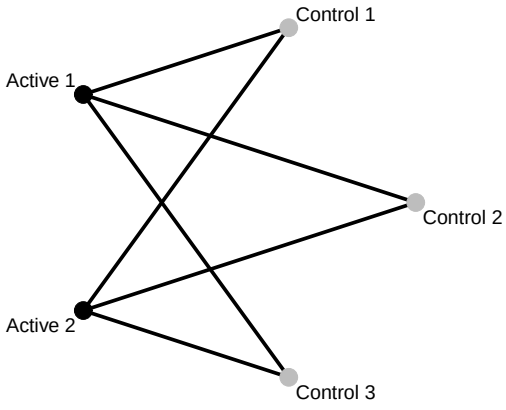

Figure 2: A multi-arm study with 2 active arms and 3 control arms.

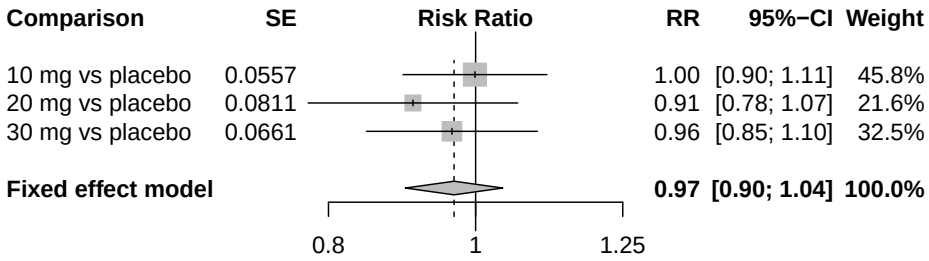

Figure 3: 'Exact adjustment' for the pelargonium sidoides data.

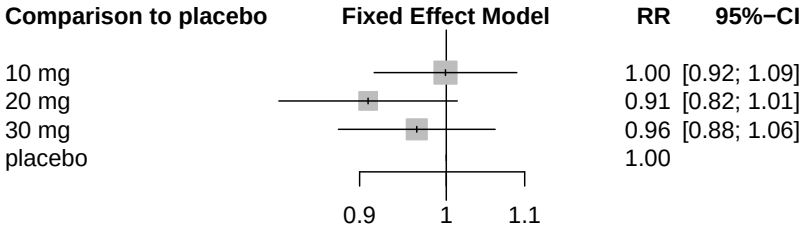

Figure 4: Network meta-analysis for the pelargonium sidoides data.

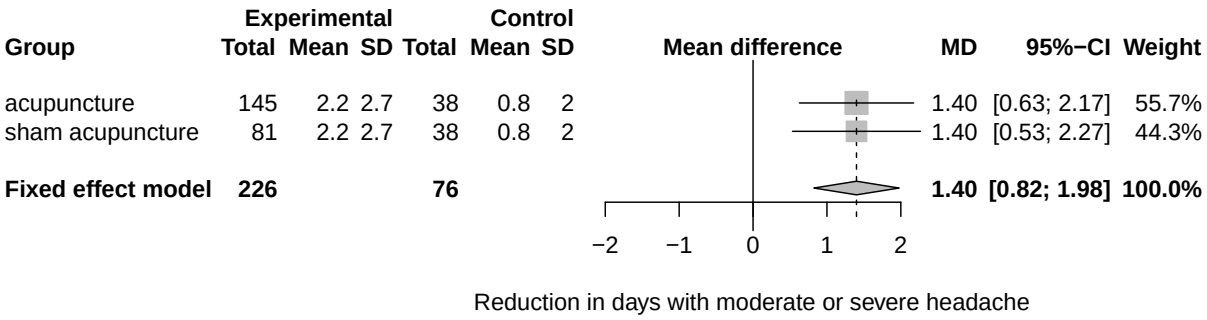

Figure 5: 'Splitting the waiting list group' for the acupuncture data.

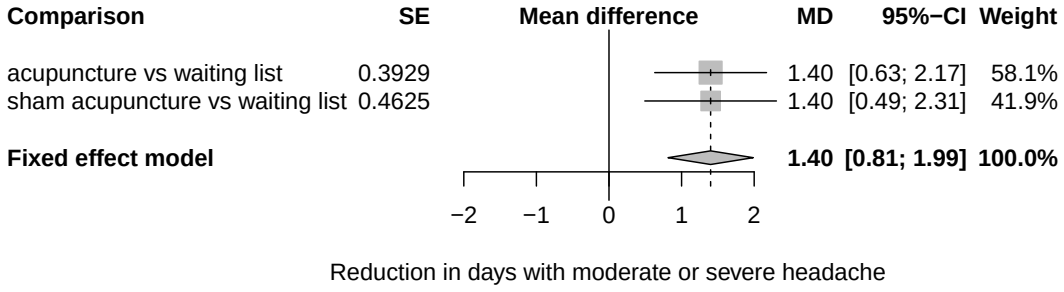

Figure 6: 'Approximate adjustment' for the acupuncture data.

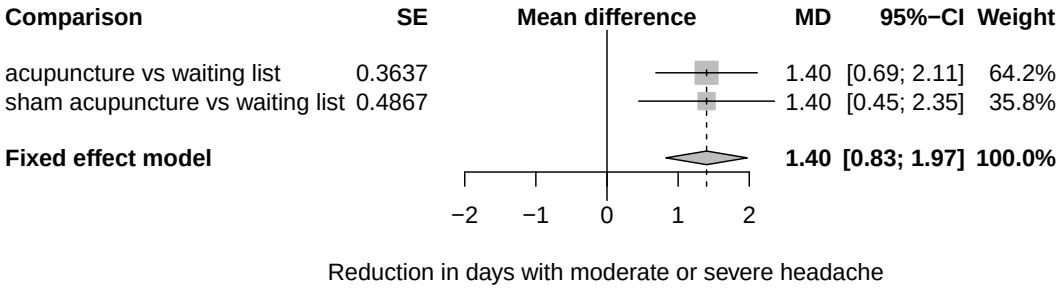

Figure 7: 'Exact adjustment' for the acupuncture data.

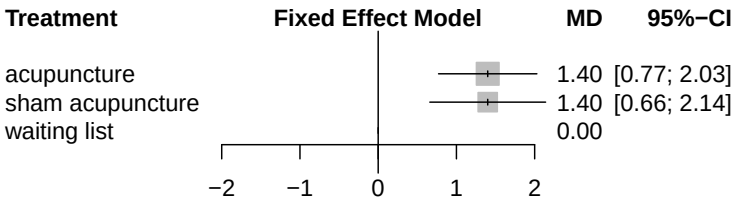

Figure 8: Network meta-analysis of the acupuncture data.

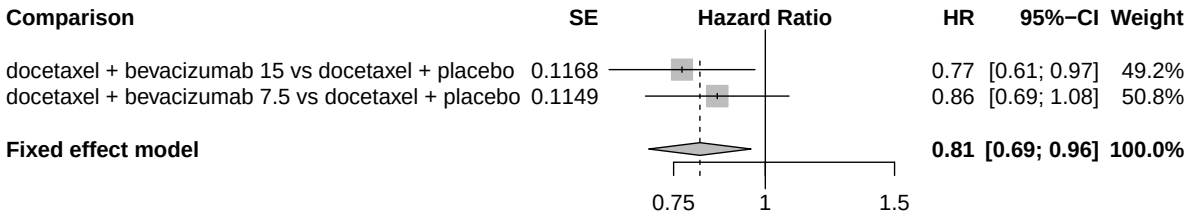

Figure 9: 'Approximate adjustment' for the breast cancer data.

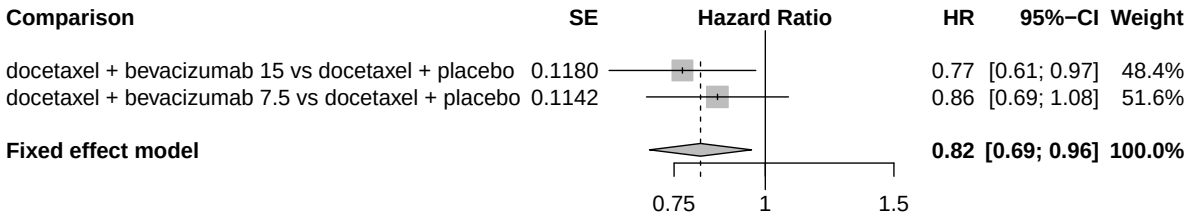

Figure 10: 'Exact adjustment' for the breast cancer data.

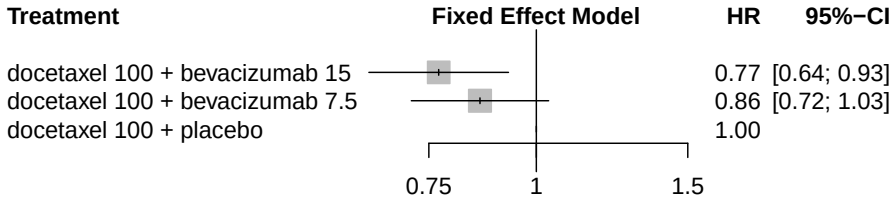

Figure 11: Network meta-analysis of the breast cancer data.