A confidence distribution approach for an efficient network meta-analysis*

Guang Yang[1], Dungang Liu[2], Regina Y. Liu[1], Minge Xie[1], and David C. Hoaglin[3]

[1]Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey 08854, U.S.A
[2]Department of Biostatistics, Yale University School of Public Health, New Haven, Connecticut 06511, U.S.A
[3]Independent Consultant, Sudbury, MA 01776, USA

Summary

This paper presents a new approach for network meta-analysis that combines multivariate confidence distributions (CDs). Network meta-analysis generalizes the traditional meta-analysis of pairwise comparisons to synthesizing studies for multiple treatment comparisons, and supports inference on all treatments in the network simultaneously. It can often strengthen inference on a pairwise comparison by borrowing evidence from other comparisons in the network. Current network meta-analysis approaches are derived from either traditional pairwise meta-analysis or hierarchical Bayesian methods. This paper introduces a general frequentist approach for network meta-analysis by combining CDs, which are viewed as frequentist "distribution estimators". Instead of combining point estimators, the proposed approach combines CD functions, which contain richer information, and thus yields greater efficiency in its inferences. This paper shows that the proposed CD approach can efficiently integrate all the studies in the network even when individual studies provide comparisons for only some of the treatments. Numerical studies, through real and simulated data sets, show that the proposed CD approach generally outperforms traditional pairwise meta-analysis and a commonly used Bayesian hierarchical model. Although the Bayesian approach may yield comparable results with a suitably chosen prior, it is sensitive to the choice of prior, which is often subjective. The CD approach is prior-free and can always provide a proper inference for the treatment effects regardless of the between-trial covariance structure.

KEY WORDS: Confidence distribution; Mixed treatment comparisons; Multiple treatment comparison; Network meta-analysis; Random-effects model.

*The research is partly supported by research grants from NSF (DMS1107012, DMS-1007683, DMS0915139, SES0851521), NSA-H98230-11-1-0157, and NIH (R01 DA016750-09)

1 Introduction

Recent advances in computing and data storage technology have greatly facilitated data gathering from many disparate sources. The demand for efficient methodologies for combining information from independent studies or disparate sources has never been greater. So far, meta-analysis is one of the most, if not the most, commonly used approaches for synthesizing findings from different sources for pairwise comparisons. For example, it is used in medical research for summarizing estimates from a set of randomized controlled trials (RCTs) of the relative efficacy of two treatments (cf. Normand, 1999; Sutton and Higgins, 2008). For more-complicated comparative effectiveness research, where the comparisons involve a network of more than two treatments, several generalizations have been developed for combining information from various sources. A useful survey can be found in the report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices (Jansen et al., 2011; Hoaglin et al., 2011) and its references. A key advantage of network meta-analysis is that it can perform indirect comparisons among multiple treatments.

We elaborate on network meta-analysis with a general setting and a worked example. In the general setting, the process begins with a systematic research for RCTs that have compared treatments for a particular condition. The trials that satisfy a set of eligibility criteria yield a network of evidence, in which each node represents a treatment and each edge represents a direct comparison in one or more trials. We assume that the network is connected, and we denote the total number of treatments by p and the number of treatments in trial i by p_i ($2 \leq p_i \leq p$). For example, Stettler et al. (2007) assembled data from 37 trials for comparing the performance of three stents in patients with coronary artery disease. Figure 1 illustrates the network of the comparisons among the three stents. Each stent is connected to the other two through a number of direct comparisons, and these three stents form a network. The primary objective is to assess the effectiveness of these three stents (more broadly all treatments in the network). The estimates of network meta-analysis yields pairwise comparisons.

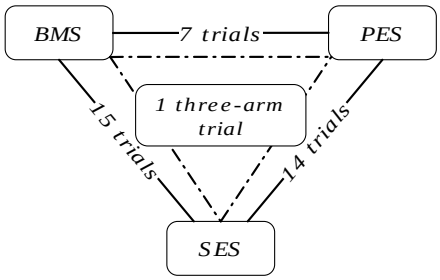

Figure 1: Network of comparisons for bare-metal stents (BMS), paclitaxel-eluting stents (PES), and sirolimus-eluting stents (SES) in 37 trials (Stettler et al., 2007)

Several network meta-analysis approaches have been reported in the literature. Lumley (2002) introduced a model for combining evidence from trials with pairwise comparisons between treatments. Although this method allows borrowing of evidence from indirect comparisons to strengthen the results of direct comparisons, it is somewhat restricted in practice because it requires that each individual trial be a two-arm trial (i.e., compare exactly two treatments).

Thus, this method cannot deal with multi-arm trials as in the example of Figure 1. Generalizing the method in Smith et al. (1995), Lu and Ades (2004) introduced a network meta-analysis approach using a Bayesian hierarchical model. Although this approach can include multi-arm trials, our simulation studies in Section 4 show that its inferences can be quite sensitive to the choice of priors. More specifically, if the assumptions in the prior distribution does not agree with the underlying true model (the unknown between-trial covariance structure), the resulting credible interval fails to achieve the nominal coverage probability, and, in some cases, its empirical coverage probability can be far below the nominal level.

This paper aims to introduce a new network meta-analysis approach that: i) can efficiently synthesize evidence from a number of independent trials on multiple treatments; ii) can include trials with multiple arms; and iii) does not need to specify priors for parameters of interest or other parameters. The proposed approach is derived from combining multivariate confidence distributions.

To some extent, our proposed CD approach extends of the method developed in Lumley (2002) to include multi-arm trials. Compared with the Bayesian method in Lu and Ades (2004), the proposed CD approach is a pure frequentist approach and it does not require specification of priors. In fact, the proposed CD approach can be viewed as a frequentist counterpart of the Bayesian method of Lu and Ades (2004).

The general idea of combining confidence distributions has been developed in Singh et al. (2005) and Xie et al. (2011). The concept of CD and its utility in statistical inference have been investigated intensely; see, e.g., Schweder and Hjort (2002) and Singh et al. (2005, 2007). A detailed survey of the recent developments on CD can be found in Xie and Singh (2013). Roughly speaking, a CD bases inferences on a sample-dependent distribution function, rather than a point or an interval, on the parameter space. A CD can be viewed as a frequentist "distribution estimator" of an unknown parameter, as described in Xie and Singh (2013) and Cox (2013). As a distribution function, a CD naturally contains more information than a point or interval estimator, and is thus a more versatile tool for inference. For example, for an odds ratio when the 2x2 table has zero events, point or interval approaches may fail, but the CD approach remains valid, as shown in Liu et al. (2012). CDs have been demonstrated in Singh et al. (2005) and Xie et al. (2011) to be especially useful for combining information on a single parameter. In particular, Xie et al. (2011) showed that the CD combining approach can provide not only a unifying framework for almost all univariate meta-analysis applications, but it can also provide new estimates that can achieve desirable properties such as high efficiency and robustness. Network meta-analysis generally involves multiple parameters, and the information on each parameter may have non-negligible impact on inferences for other parameters. To fully utilize the joint information on multiple parameters, we construct multivariate joint CD functions for the entire set of parameters from each study. The combination of these joint CD functions leads to a novel frequentist approach to network meta-analysis.

Our numerical studies show that the proposed CD approach compares favorably with, and often is superior to, traditional meta-analysis and the hierarchical Bayesian network meta-analysis method proposed by Lu and Ades (2004). Specifically, in comparison with the traditional method, the CD method is more efficient because it uses indirect evidence. In comparison with the Bayesian method, the CD approach is prior-free and can always provide a proper

inference (i.e., confidence intervals with correct coverage rates) for treatment effects, regardless of the between-trial covariance structure. Moreover, our simulation studies show that the performance of the Bayesian approach is sensitive to the choice of prior distributions, which ideally should reflect the true underlying the between-trial covariance structure.

The paper is organized as follows. Section 2 reviews the concept of CD and develops a general method for combining multivariate normal CDs to facilitate network meta-analysis. Section 3 uses two real data examples to illustrate the proposed CD approach in the analysis of a three-treatment network, and to compare it with traditional meta-analysis and the Bayesian network meta-analysis. In Section 4, the results of several simulation studies demonstrate that the proposed CD approach can provide proper inferences. Comparisons with the traditional and Bayesian network meta-analysis approaches are also provided. Moreover, we devise a simple adaptive CD approach to address possible inconsistent (or contradictory) evidence from indirect and direct comparisons. This adaptive approach can alleviate undue influence from indirect comparisons whose evidence contradicts the direct comparisons. Section 5 contains a summary and further remarks.

2 A CD approach for network meta-analysis

Assume that the evidence network comprises k independent clinical trials and involves the effects of p treatments, denoted by the vector $\boldsymbol{\theta} \equiv (\theta_1, \cdots, \theta_p)^{\mathbf{T}}$. The individual trials may have studied only a subset of the p treatments. More specifically, the i-th trial involves $p_i \leq p$ treatments. If $p_i < p$, the i-th trial provides only partial information about $\boldsymbol{\theta}$, in the sense that only the p_i-dimensional parameter $\boldsymbol{\theta}_i \equiv \mathbf{A}_i\boldsymbol{\theta}$ is identifiable, where the $p_i \times p$ selection matrix $\mathbf{A}_i$ is obtained by removing from the $p \times p$ identity matrix (or, more generally, any $p \times p$ orthogonal matrix $\mathbf{A}$) the rows that correspond to the omitted parameters. Throughout this paper, we consider the following multivariate random-effects model for network meta-analysis. It extends the univariate hierarchical random-effects model reviewed in Normand (1999):

$$\boldsymbol{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i \overset{\text{ind}}{\sim} N(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i), \quad \boldsymbol{\theta}_i|\boldsymbol{\theta}, \mathbf{S} \overset{\text{ind}}{\sim} N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}), \quad i = 1, 2, \ldots, k \tag{1}$$

where $\boldsymbol{y}_i$ is the summary statistic from the i-th study, $\boldsymbol{\Sigma}_i$ is the covariance matrix of $\boldsymbol{y}_i$, and $\mathbf{S}$ is red the covariance matrix of random-effects distribution.

A key question in network meta-analysis is how the information on $\boldsymbol{\theta}_i$ (which may provide only partial information on $\boldsymbol{\theta}$) can be integrated to make efficient inference about $\boldsymbol{\theta}$. Our proposed approach of combining multivariate normal CDs can provide a solution.

Before presenting our CD approach for network meta-analysis, we review the combining CD procedure for the univariate case in Section 2.1 and then extend it to the multivariate case in Section 2.2.

2.1 Review of CD approach for univariate meta-analysis

We first consider the special case where the parameter of interest is univariate. Model (1) simplifies to model (2)-(3) of Normand (1999); i.e.,

$$y_i|\theta_i, \sigma_i^2 \overset{\text{ind}}{\sim} N(\theta_i, \sigma_i^2), \quad \theta_i|\theta, \tau^2 \overset{\text{ind}}{\sim} N(\theta, \tau^2), \quad i = 1, 2, \ldots, k \tag{2}$$

where θ_i is the study-specific mean (random-effect) and θ and τ^2 are hyper-parameters for θ_i.

For the univariate case, meta-analysis estimators used in current practice (c.f., Table IV of Normand, 1999) can all be obtained through the unifying framework developed by Xie et al. (2011) using the CD concept. A CD has been loosely referred to as a distribution function on the parameter space that can represent confidence intervals of all levels for a given parameter of interest. More specifically, the following formal definition is proposed in Schweder and Hjort (2002) and Singh et al. (2005, 2007):

Definition 1 *Suppose Θ is the parameter space of the unknown parameter of interest θ, and $\mathcal{Y}$ is the sample space corresponding to data $\mathbf{Y} = \{y_1, \ldots, y_n\}$. Then a function $H(\cdot) = H(\mathbf{Y}, \cdot)$ on $\mathcal{Y} \times \Theta \to [0, 1]$ is a confidence distribution (CD) if:*
(i) For each given $\mathbf{Y} \in \mathcal{Y}$, $H(\cdot)$ is a continuous cumulative distribution function on Θ; and
(ii) At the true parameter value $\theta = \theta_0$, $H(\theta_0) = H(\mathbf{Y}, \theta_0)$, as a function of the sample $\mathbf{Y}$, follows the uniform distribution $U[0, 1]$.
The function $H(\cdot)$ is an asymptotic CD (aCD) if the $U[0, 1]$ requirement holds only asymptotically and the continuity requirement on $H(\cdot)$ is dropped.

In other words, a confidence distribution is a function defined on both the parameter space and the sample space, satisfying requirements (i) and (ii). Requirement (i) simply says that a CD should be a distribution on the parameter space. Requirement (ii) imposes some restrictions to facilitate desirable frequentist properties such as unbiasedness, consistency and/or efficiency. The CD concept is broad, covering examples from regular parametric (fiducial distribution) to bootstrap distributions, significance functions (also called p-value functions), normalized likelihood functions, and, in some cases, Bayesian priors and posteriors; see, e.g., Singh et al. (2007) and Xie and Singh (2013). A CD can be used to draw various inferences for the unknown parameter. For example, the median/mean of the distribution function $H(\cdot)$ can be used as a point estimator of θ, and the interval $(-\infty, H^{-1}(1-\alpha))$ forms a level $(1-\alpha)$ confidence interval, an immediate consequence of Requirement (ii).

Example 1 (*CDs for univariate normal mean*) Let $\{y_i, i = 1, \ldots, n\}$ be an iid sample from $N(\theta, \sigma^2)$ with mean $\bar{y}$. Suppose that the parameter θ is of primary interest. If σ^2 is known, then $H_\Phi(\theta) = \Phi(\sqrt{n}(\theta - \bar{y})/\sigma)$ satisfies the two requirements in Definition 1, and it is a CD for θ. If σ^2 is unknown, one can show that $H_t(\theta) = F_{t_{n-1}}(\sqrt{n}(\theta - \bar{y})/s)$ is a CD for θ. Here s^2 is the sample variance, and $F_{t_{n-1}}$ is the cumulative distribution function of the student-t distribution with $(n-1)$ degrees of freedom. However, $H_A(\theta) = \Phi(\sqrt{n}(\theta - \bar{y})/s)$ is only an asymptotic CD for θ.

To combine individual CDs $H_i(\theta) = H_i(\boldsymbol{y}_i, \theta), i = 1, \ldots, k$, Singh et al. (2005) proposed a general recipe that uses a coordinate-wise monotonic function that maps the k-dimensional

cube $[0,1]^k$ to the real line. Specifically, a combined CD can be constructed following

$$H^{(c)}(\theta) = G^{(c)}\{g^{(c)}(H_1(\theta), \ldots, H_k(\theta))\}, \tag{3}$$

where the function $G^{(c)}$ is defined as $G^{(c)}(t) = \Pr\{g^{(c)}(U_1, \ldots, U_k) \leq t\}$ in which $U_1, \ldots, U_k$ are independent $U[0,1]$ random variables. Xie et al. (2011) applied this general recipe to meta-analysis, with a special choice of $g^{(c)}$:

$$g^{(c)}(u_1, \ldots, u_k) = \widetilde{w}_1 a_0(u_1) + \cdots + \widetilde{w}_k a_0(u_k), \tag{4}$$

where $a_0(\cdot)$ is a given monotonic function and $\widetilde{w}_i \geq 0$, with at least one $\widetilde{w}_i \neq 0$, are generic weights for the combination. Xie et al. (2011) and subsequent research showed that, with suitable choices of $g^{(c)}$, almost all combining methods currently used in meta-analysis can be unified under the framework of Equation (3), including p-value combination methods, model-based meta-analysis (fixed-effect and random-effects models), the Mantel-Haenszel method, Peto's method, and also the method in Tian et al. (2009) by combining confidence intervals.

For the special model in (2), one can construct $H_i(\theta) = \Phi((\theta - y_i)/(\sigma_i^2 + \tau^2)^{1/2})$ based on the ith study and take $a_0(\cdot) = \Phi^{-1}(\cdot)$ and $\widetilde{w}_i = 1/(\sigma_i^2 + \tau^2)^{1/2}$ in (4). Here τ^2 is assumed known. If τ^2 is unknown, one can replace it with the DerSimonian and Laird estimator $\widehat{\tau}^2_{\texttt{DL}}$ (DerSimonian and Laird, 1986) or preferably the restricted-maximum-likelihood estimator $\widehat{\tau}^2_{\texttt{REML}}$. Then the combined CD function for θ is

$$H^{(c)}(\theta) = \Phi\left(\left(\sum_{i=1}^{k} \frac{1}{\sigma_i^2 + \tau^2}\right)^{1/2} (\theta - \widehat{\theta}^{(c)})\right), \tag{5}$$

where $\widehat{\theta}^{(c)} = \{\sum_{i=1}^k \frac{y_i}{\sigma_i^2+\tau^2}\}/\{\sum_{i=1}^k \frac{1}{\sigma_i^2+\tau^2}\}$. The combined CD function is normal with mean $\widehat{\theta}^{(c)}$ and variance $s_c^2 = \{\sum_{i=1}^k \frac{1}{\sigma_i^2+\tau^2}\}^{-1}$, which is ready for making point estimates and constructing confidence intervals for the parameter θ.

From Definition 1, a CD function $H(\cdot)$ is a cumulative distribution function on the parameter space for each given sample $\mathbf{Y}_n$. Thus, we can construct a random variable ξ defined on $\mathcal{Y} \times \Theta$ such that, conditional on the sample, ξ has the distribution $H(\cdot)$. We call this random variable ξ a *CD random variable* (see, e.g., Singh et al., 2007; Xie and Singh, 2013). Conversely, suppose we have a CD random variable $\xi \in \mathcal{Y} \times \Theta$ whose conditional distribution, conditional on the sample, has a cumulative distribution function $H(\cdot)$. Then $H(\cdot)$ is a CD for the parameter of interest θ.

We can express the normal CD combination (5) as a combination of CD random variables. Specifically, for a CD-random variable $\xi_i | y_i \sim H_i(\theta) = \Phi((\theta - y_i)/(\sigma_i^2 + \tau^2)^{1/2})$ derived from the i-th study, we can define $\xi^{(c)} = \sum_{i=1}^k w_i \xi_i$, where $w_i = 1/(\sigma_i^2 + \tau^2)$, and its corresponding combined CD

$$H^{(c)}(\theta) = \Pr(\xi^{(c)} \leq \theta | data), \quad \text{for any } \theta \in \Theta. \tag{6}$$

It is straightforward to show that the $H^{(c)}(\cdot)$ defined in (6) is the same as the one defined in (5).

The concept of CD random variable has been investigated in several recent publications. Xie and Singh (2013) explored the connection of CD random variables with bootstrap estimators when the bootstrap approach applies. Hannig and Xie (2012) discussed the association of a CD random variable with the so-called *belief random set*, a fundamental concept in the Dempster-Shafer theory of belief functions (cf. Dempster, 2008; Martin and Liu, 2013).

2.2 A general procedure to combine multivariate normal CDs

Constructing and combining CDs for multi-dimensional parameters is not a straightforward extension of the univariate case. One difficulty is that the cumulative distribution function is not a useful notion in the multivariate case, because (a) the region $F(\mathbf{y}) \leq \alpha$ is not of main interest and (b) the property $F(\boldsymbol{Y}) \overset{L}{=} U[0,1]$ when $\boldsymbol{Y} \overset{L}{=} F$ does not hold in $\Re^p$ (Singh et al., 2007). Research thus far suggests that we either limit our interest to center-outward confidence regions (instead of all Borel sets) in the $p \times 1$ parameter space or use asymptotic normality; see Xie and Singh (2013) and also De Blasi and Schweder (2012). In the present context, it suffices to consider only the multivariate normal CDs because individual CDs are based on asymptotic normality. We use a multivariate normal CD definition proposed in Singh et al. (2007). Intuitively, a distribution function $H(\cdot)$ is a multivariate normal CD for a $p \times 1$ vector $\boldsymbol{\theta}$ if and only if the projected distribution of $H(\cdot)$ on any direction $\boldsymbol{\lambda} \in \Re^p$, $||\boldsymbol{\lambda}||_2 = 1$, is a univariate normal CD for $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$. Here is a formal definition of a multivariate normal CD:

Definition 2 *Let $\boldsymbol{\xi}$ be a random vector on $\Re^p$. For any given $p \times 1$ vector $\boldsymbol{\lambda}$, $||\boldsymbol{\lambda}||_2 = 1$, we denote by $H_{\boldsymbol{\lambda}}(\cdot)$ the conditional distribution of $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\xi}$ given $\mathbf{Y}$. We also denote by $H(\cdot)$ the conditional distribution of $\boldsymbol{\xi}$ given $\mathbf{Y}$. Then we call $H(\cdot)$ the multivariate normal CD (or, asymptotic multivariate normal CD) for a $p \times 1$ parameter vector $\boldsymbol{\theta}$ if and only if, for any given $\boldsymbol{\lambda}$, $H_{\boldsymbol{\lambda}}(\cdot)$ is a univariate normal CD (or asymptotic CD) function for $\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\theta}$. Also, the random vector $\boldsymbol{\xi}$ is called a CD random vector for $\boldsymbol{\theta}$.*

Example 2 (*CDs for multivariate normal mean*) Suppose $\boldsymbol{y}_i, i = 1, \ldots, n$ are identically and independently distributed observations from a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}$ is known, then the sample-dependent distribution $N(\bar{\boldsymbol{y}}, \boldsymbol{\Sigma})$ is a multivariate normal CD function for $\boldsymbol{\theta}$, where $\bar{\boldsymbol{y}}$ is the sample mean. If $\boldsymbol{\Sigma}$ is unknown but can be estimated consistently, say by $\widehat{\boldsymbol{\Sigma}}$, then the sample-dependent distribution $N(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Sigma}})$ is an asymptotic multivariate normal CD function for $\boldsymbol{\theta}$.

The CD combination method for the multivariate case cannot be easily specified by following (3) and (4), especially under the setting of (1), where p_i may differ. Instead, we utilize the concept of CD random vector and an extension of (6) to propose the following scheme for combining multivariate normal CDs.

Theorem 1 *Let $H_i(\boldsymbol{\theta}_i) \equiv H_i(\mathbf{Y}_i, \boldsymbol{\theta}_i), i = 1 \ldots, k$ are multivariate normal CD functions for the multivariate parameters $\boldsymbol{\theta}_i$ from k independent samples $\mathbf{Y}_i$, where $\boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\theta}$ for the same p-dimensional target parameter vector $\boldsymbol{\theta}$. Additionally, let $\boldsymbol{\xi}_i$ be the CD random vector for $\boldsymbol{\theta}_i$.*

For any $\mathbf{t} \in \Re^p$*, we define*

$$H^{(c)}(\mathbf{t}) = \Pr\left\{ \left(\sum_{i=1}^{k} W_i \right)^{-1} \sum_{i=1}^{k} W_i \mathbf{A}_i^{+} \boldsymbol{\xi}_i \leq \mathbf{t} \middle| \mathbf{Y}_1, \ldots, \mathbf{Y}_k \right\}, \tag{7}$$

where $\mathbf{A}_i^{+}$ *is the Moore-Penrose pseudo-inverse of* $\mathbf{A}_i$*. Then* $H^{(c)}(\cdot) = H(\mathbf{Y}_1, \ldots, \mathbf{Y}_k; \cdot)$ *is a multivariate normal CD for* $\boldsymbol{\theta}$ *provided the following conditions hold:*
(1) Each $p \times p$ *matrix* W_i *is positive semi-definite.*
(2) $\mathscr{C}(W_i) = V_i$*, where* $\mathscr{C}(W_i)$ *is the column space of* W_i *and* V_i *is the row space of* $\mathbf{A}_i$*.*
(3) $V_1 + V_2 + \cdots + V_k = \Re^p$*, where* $V_1 + V_2 + \cdots + V_k \overset{\triangle}{=} \{\sum_{i=1}^{k} v_i | v_i \in V_i, i = 1, \ldots, k\}$*.*

In Theorem 1, conditions (2) and (3) state that, even if $\text{rank}(\mathbf{A}_i) < p$ for all i, so that $\boldsymbol{\theta}$ is not identifiable in any individual study, we can still derive a multivariate normal CD for $\boldsymbol{\theta}$ as long as the treatments are connected in a network.

Recall the multivariate model introduced in (1). We first consider the case in which $\boldsymbol{\Sigma}_i$ and $\mathbf{S}$ are known. From Example 2, we know that $N(\boldsymbol{y}_i, \boldsymbol{\Sigma}_i + \mathbf{A}_i \mathbf{S} \mathbf{A}_i^{\mathbf{T}})$ is a multivariate normal CD function for $\boldsymbol{\theta}_i$ based on the i-th study. Let $\boldsymbol{\xi}_i$ be the corresponding CD random vector for inference on $\boldsymbol{\theta}_i$ and $W_i = \mathbf{A}_i^{+}(\boldsymbol{\Sigma}_i + \mathbf{A}_i \mathbf{S} \mathbf{A}_i^{\mathbf{T}})^{-1} \mathbf{A}_i$. It follows that $\left(\sum_{i=1}^{k} W_i\right)^{-1} \sum_{i=1}^{k} W_i \mathbf{A}_i^{+} \boldsymbol{\xi}_i$ is normally distributed with mean vector $\widehat{\boldsymbol{\theta}}^{(c)} = (\sum_{i=1}^{k} W_i)^{-1} (\sum_{i=1}^{k} W_i \mathbf{A}_i^{+} \boldsymbol{y}_i)$ and variance $\mathbf{S}_c = (\sum_{i=1}^{k} W_i)^{-1}$, given the sample. Thus, following the recipe in Equation (7), the combined CD for $\boldsymbol{\theta}$ is

$$H^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\left(\mathbf{S}_c^{-1/2} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(c)}) \right) \tag{8}$$

where $\boldsymbol{\Psi}(\cdot)$ is the cdf of the standard $p \times 1$ multivariate normal distribution function. Conditions (1) and (2) of Theorem 1 are satisfied by the specification of W_i, and condition (3) is satisfied as long as the comparisons involved in the studies form a connected network. Based on the combined multivariate CD function in (8), we can use $\widehat{\boldsymbol{\theta}}^{(c)}$ as a point estimator for $\boldsymbol{\theta}$ with variance $\mathbf{S}_c$. Furthermore, inferences on any linear contrasts $\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\theta}$ of $\boldsymbol{\theta}$ can be obtained from $\boldsymbol{\lambda}^{\mathbf{T}} \boldsymbol{\xi}^{(c)}$, where $\boldsymbol{\xi}^{(c)}$ follows the distribution specified in Equation (8).

If $\boldsymbol{\Sigma}_i$ and $\mathbf{S}$ are unknown, we can replace them with the sample estimators $\widehat{\boldsymbol{\Sigma}}_i$ and $\mathbf{S}_{\texttt{REML}}$. Then, as long as these estimators are consistent, the distribution $N(\boldsymbol{y}_i, \widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i \boldsymbol{S}_{\texttt{REML}} \mathbf{A}_i^{\mathbf{T}})$ is asymptotically a multivariate normal CD for $\boldsymbol{\theta}_i$. Here $\widehat{\boldsymbol{\Sigma}}_i$ is the sample covariance matrix, and $\mathbf{S}_{\texttt{REML}}$ is the restricted-maximum-likelihood estimator of the heterogeneity between studies. As a result, the combined CD function (8) is an asymptotic multivariate normal CD for $\boldsymbol{\theta}$ with $\boldsymbol{\Sigma}_i$ and $\mathbf{S}$ replaced by $\widehat{\boldsymbol{\Sigma}}_i$ and $\mathbf{S}_{\texttt{REML}}$, respectively. For the estimation of $\mathbf{S}$, Jackson et al. (2010) developed a direct extension of the DerSimonian and Laird estimator of heterogeneity to multivariate case. Hereafter, we denote by $\mathbf{S}_{\texttt{DL}}$ and $\mathbf{S}_{\texttt{REML}}$ respectively the estimator derived from Jackson et al. (2010) and the restricted-maximum-likelihood estimator. We apply and examine both estimators in our numerical study of real examples and simulations in Sections 3 and 4. Further discussions on the performance of the DL and REML estimators for the heterogeneity in univariate random-effects models can be found in Sidik and Jonkman (2007) and Thorlund et al. (2011).

3 Real data examples

In this section, we illustrate the proposed CD approach for network meta-analysis using two real data examples, one on coronary artery disease and the other on cirrhosis. For comparison, we also include the traditional pairwise meta-analysis and the Bayesian hierarchical model.

3.1 An example on coronary artery disease (CAD)

Stettler et al. (2007) used data from a network of 37 trials to compare the performance of three types of stent: bare metal stent (BMS), sirolimus-eluting stent (SES), and paclitaxel-eluting stent (PES), in patients with coronary artery disease. Each trial involved at least two of the three treatments; we analyze the data on a negative outcome, whether patients required target lesion revascularisation (TLR) within one year (cf. Figure 1). One trial, TAXUS I, had zero events and is thus excluded from the analysis. Of the remaining 36 trials, listed in Table 1, 15 trials compared BMS with SES, 6 trials compared BMS with PES, 14 trials compared SES with PES, and 1 trial compared all three treatments. The network is connected, so simultaneous inference on the treatment effects is possible.

3.1.1 A multivariate random-effects model

We use treatments A, B, and C to denote the three types of stents BMS, SES and PES, respectively. We use T_i to denote the set of treatments compared in the i-th trial; for example, $T_i = \{A, C\}$ for TAXUS IV. Further, let n_{ij} and r_{ij} be the number of total patients and number of patients who experienced a TLR in the i-th study with treatment j. Then with a binary individual responses we would assume

$$r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad i = 1, 2, \ldots, 36, j \in T_i \tag{9}$$

where p_{ij} denotes the probability that a patient on treatment j experiences an event in the i-th trial.

The target parameter is $\mathbf{p} = (p_A, p_B, p_C)^{\mathbf{T}}$, the overall probability of an event for BMS, SES, and PES, respectively. In practice, one often applies a log transformation to the observed odds of an event. Owing to the rapid convergence to a normal distribution on the log-odds scale, it is customary to consider a general random-effects model for $\boldsymbol{\theta}_i = (\text{logit}(p_{ij}))^{\mathbf{T}}, \forall j \in T_i$ with parameter $\boldsymbol{\theta} = (\text{logit}(p_A), \text{logit}(p_B), \text{logit}(p_C))^{\mathbf{T}}$; cf. DerSimonian and Laird (1986); Normand (1999). Here, $\text{logit}(p) = \log(p/(1-p))$. Specifically, we have

$$\begin{array}{ll} \text{level 1:} & r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad i = 1, 2, \ldots, 36, j \in T_i \\ \text{level 2:} & \boldsymbol{\theta}_i \sim N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}) \end{array} \tag{10}$$

where $\mathbf{A}_i$ is the selection matrix associated with T_i; for example, $\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ if $T_i = \{A, B\}$, $\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ if $T_i = \{A, C\}$, $\mathbf{A}_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ if $T_i = \{B, C\}$, and $\mathbf{A}_i = \mathbf{I}_3$ if $T_i = \{A, B, C\}$.

Table 1: CAD Trial Data, Target Lesion Revascularisation at 1 year

Study	BMS (A)		SES (B)		PES (C)	
	r_{ij}	n_{ij}	r_{ij}	n_{ij}	r_{ij}	n_{ij}
BASKET	35	281	25	264	25	281
C-SIRIUS	11	50	2	50	—	—
DECODE	8	29	5	54	—	—
DIABETES	27	80	6	80	—	—
E-SIRIUS	44	177	8	175	—	—
Ortolani 2007	11	52	6	52	—	—
Pache 2005	51	250	25	250	—	—
PRISON II	20	100	4	100	—	—
RAVEL	16	118	1	120	—	—
RRISC	10	37	6	38	—	—
SCANDSTENT	47	159	4	163	—	—
SCORPIUS	20	95	5	95	—	—
SESAMI	19	160	7	160	—	—
SES-SMART	27	128	9	129	—	—
SIRIUS	106	525	26	533	—	—
TYPHOON	45	357	13	355	—	—
HAAMUS-TENT	9	82	—	—	3	82
PASSION	23	309	—	—	16	310
TAXUS II	39	269	—	—	13	260
TAXUS IV	96	652	—	—	28	662
TAXUS V	107	579	—	—	62	577
TAXUS VI	46	227	—	—	19	219
Cervinka 2006	—	—	1	37	2	33
CORPAL	—	—	22	331	25	321
Han 2006	—	—	9	202	11	196
ISAR-DESIRE	—	—	14	100	22	100
ISAR-DIABETES	—	—	9	125	15	125
ISAR-SMART3	—	—	16	180	29	180
LONG DES II	—	—	6	250	18	250
Petronio 2007	—	—	1	42	1	43
PROSIT	—	—	3	116	9	115
REALITY	—	—	44	684	43	669
SIRTAX	—	—	30	503	54	509
SORT OUT II	—	—	40	1065	46	1033
TAXi	—	—	4	102	2	100
Zhang 2006	—	—	14	225	16	187

Further, let $y_{ij} = \log\left(\frac{r_{ij}}{n_{ij}-r_{ij}}\right), \widehat{\sigma}^2_{ij} = \frac{1}{r_{ij}} + \frac{1}{n_{ij}-r_{ij}}$ and $\boldsymbol{y}_i = [y_{ij}, j \in T_i]^{\mathbf{T}}, \widehat{\boldsymbol{\Sigma}}_i =$

diag$(\widehat{\sigma}_{ij}^2, j \in T_i)$. Then an asymptotically equivalent model is

$$\begin{array}{ll} \text{level 1:} & \boldsymbol{y}_i|\boldsymbol{\theta}_i \sim N(\boldsymbol{\theta}_i, \widehat{\boldsymbol{\Sigma}}_i), \quad i = 1, 2, \ldots, 36 \\ \text{level 2:} & \boldsymbol{\theta}_i \sim N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}). \end{array} \tag{11}$$

Finally, if that our primary concern is the efficacy of SES vs BMS, the parameter of interest is the log-odds ratio reflecting the relative efficacy of treatment B vs A, that is $\delta_{AB} \equiv \theta_B - \theta_A$. We proceed to compare the results obtained from the proposed CD procedure with those from the traditional pairwise meta-analysis and the Bayesian network meta-analysis.

3.1.2 The CD approach

Consider the random-effects model in (11). We estimate the covariance matrix $\mathbf{S}$ by the restricted-maximum-likelihood estimator $\mathbf{S}_{\texttt{REML}}$. We can construct a multivariate normal aCD function for $\boldsymbol{\theta}_i$ based on the i-th individual study, namely $N(\boldsymbol{y}_i, \widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i\mathbf{S}_{\texttt{REML}}\mathbf{A}_i^{\mathbf{T}})$. We use $\boldsymbol{\xi}_i|\boldsymbol{y}_i \sim N(\boldsymbol{y}_i, \widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i\mathbf{S}_{\texttt{REML}}\mathbf{A}_i^{\mathbf{T}})$ to denote the associated CD random variable and take $W_i = \mathbf{A}_i^{+}(\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i\mathbf{S}_{\texttt{REML}}\mathbf{A}_i^{\mathbf{T}})^{-1}\mathbf{A}_i$. Then, by (8), $H^{(c)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\mathbf{S}_c^{-1/2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(c)}))$ is the combined CD for $\boldsymbol{\theta}$, where $\widehat{\boldsymbol{\theta}}^{(c)} = (\sum_{i=1}^k W_i)^{-1}(\sum_{i=1}^k W_i\mathbf{A}_i^{+}\boldsymbol{y}_i)$ and $\mathbf{S}_c = (\sum_{i=1}^k W_i)^{-1}$. Since we have $\mathbf{A}_i^{+} = \mathbf{A}_i^{\mathbf{T}}$ in the current case, we can replace $\mathbf{A}_i^{+}$ with $\mathbf{A}_i^{\mathbf{T}}$ in the above formulas.

To make inferences for $\delta_{AB} \equiv \theta_B - \theta_A$, we can use the marginal distribution of $\boldsymbol{\lambda}_{AB}^{\mathbf{T}}\boldsymbol{\xi}^{(c)}$ where $\boldsymbol{\lambda}_{AB} = (-1, 1, 0)^{\mathbf{T}}$ and $\boldsymbol{\xi}^{(c)}|data \sim N(\widehat{\boldsymbol{\theta}}^{(c)}, \mathbf{S}_c)$. Therefore, the point estimator $\widehat{\delta}_{AB}$ and its variance based on the CD procedure are

$$\widehat{\delta}_{AB} = \boldsymbol{\lambda}_{AB}^{\mathbf{T}}\left\{\sum_{i=1}^k \mathbf{A}_i^{+}(\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i\mathbf{S}_{\texttt{REML}}\mathbf{A}_i^{\mathbf{T}})^{-1}\mathbf{A}_i\right\}^{-1}\sum_{i=1}^k \mathbf{A}_i^{+}(\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i\mathbf{S}_{\texttt{REML}}\mathbf{A}_i^{\mathbf{T}})^{-1}\mathbf{A}_i\mathbf{A}_i^{\mathbf{T}}\boldsymbol{y}_i$$

$$\text{var}(\widehat{\delta}_{AB}) = \boldsymbol{\lambda}_{AB}^{\mathbf{T}}\left\{\sum_{i=1}^k \mathbf{A}_i^{+}(\widehat{\boldsymbol{\Sigma}}_i + \mathbf{A}_i\mathbf{S}_{\texttt{REML}}\mathbf{A}_i^{\mathbf{T}})^{-1}\mathbf{A}_i\right\}^{-1}\boldsymbol{\lambda}_{AB}.$$

In practice, we might also be interested in simultaneous inferences on, say, q linear combinations of $\boldsymbol{\theta}$, e.g., $\mathbf{Q}\boldsymbol{\theta}$ where $\mathbf{Q} \in \Re^{q\times p}$. The Bayesian approach often uses the marginal posterior distribution of $\mathbf{Q}\boldsymbol{\theta}$ as the basis for statistical inference. Similarly, to draw inferences for $\boldsymbol{\theta}$, the proposed CD network meta-analysis approach can use the marginal distribution of $\mathbf{Q}\boldsymbol{\xi}^{(c)}$ given the data. Here $\boldsymbol{\xi}^{(c)}$ is the CD random vector associated with the combined CD function $H^{(c)}(\cdot)$ for $\boldsymbol{\theta}$.

3.1.3 Traditional pairwise meta-analysis

A traditional meta-analysis for such a problem uses only the direct evidence, e.g., clinical trials that explicitly compared BMS vs SES; see, e.g., Simmonds and Higgins (2007) and Hoaglin

et al. (2011). Let $\widehat{\delta}_{AB,i} = \log\left(\dfrac{r_{iB}(n_{iA}-r_{iA})}{r_{iA}(n_{iB}-r_{iB})}\right)$ for $A, B \in T_i$. A random-effects model (DerSimonian and Laird, 1986) is considered:

$$\begin{array}{ll} \text{level 1:} & \widehat{\delta}_{AB,i} \sim N(\delta_{AB,i}, \sigma^2_{AB,i}), \quad i \text{ s.t. } A, B \in T_i \\ \text{level 2:} & \delta_{AB,i} \sim N(\delta_{AB}, \tau^2_{AB}). \end{array} \tag{12}$$

An overall estimate of the common log-odds ratio δ_{AB}, based on the direct evidence, is often a weighted average of the estimates $\widehat{\delta}_{AB,i}$ from individual studies (Hardy and Thompson, 1996):

$$\widehat{\delta}_{AB,direct} = \frac{\sum_i w_i \widehat{\delta}_{AB,i}}{\sum_i w_i} \quad \text{with} \quad \text{var}(\widehat{\delta}_{AB}) = \frac{1}{\sum_i w_i}, \tag{13}$$

where the weight w_i is often taken as the empirical weight determined by the reciprocal of the variance $\sigma^2_{AB,i}$ adjusted to incorporate the heterogeneity τ^2_{AB}, for example $w_i = 1/(\sigma^2_{AB,i}+\tau^2_{AB})$, as suggested in DerSimonian and Laird (1986).

In practice, when the variance $\sigma^2_{AB,i}$ and the heterogeneity τ^2_{AB} are unknown, they are often replaced by their corresponding estimates $\widehat{\sigma}^2_{AB,i}$ and $\widehat{\tau}^2_{AB}$, where $\widehat{\sigma}^2_{AB,i} = \frac{1}{r_{iA}} + \frac{1}{n_{iA}-r_{iA}} + \frac{1}{r_{iB}} + \frac{1}{n_{iB}-r_{iB}}$, provided that $r_{ij} \neq 0$ and $r_{ij} \neq n_{ij}$, and $\widehat{\tau}^2_{AB}$ is the REML estimate.

Similarly, we can obtain estimates $\widehat{\delta}_{AC}$ and $\widehat{\delta}_{BC}$ for the pairwise comparisons of BMS vs PES and SES vs PES, respectively, based on the 7 and 15 trials that compared them directly. Then an indirect comparison of BMS vs SES can be obtained by taking

$$\widehat{\delta}_{AB,indirect} = \widehat{\delta}_{AC} - \widehat{\delta}_{BC} \quad \text{and} \quad \text{var}(\widehat{\delta}_{AB}) = \text{var}(\widehat{\delta}_{AC}) + \text{var}(\widehat{\delta}_{BC}). \tag{14}$$

We can then combine the $\widehat{\delta}_{AB,direct}$ and $\widehat{\delta}_{AB,indirect}$ to obtain an estimator that integrates the two sources of information, provided that the direct and indirect comparisons are consistent with each other or at least not contradictory. Here is a simple illustration of inconsistent/contradictory evidence: the direct comparison concludes that the effect of treatment X is larger than that of treatment Y, but the indirect comparison concludes the opposite. Some discussion on issues of inconsistent evidence in network meta-analysis can be found in Lumley (2002), Lu and Ades (2006), and Dias et al. (2010).

Although one can always apply the procedure above to combine the direct and indirect estimates, this procedure splits the three-arm trial into three two-arm trials and uses them for three difference estimates. This is a drawback for traditional pairwise meta-analysis — *Trials with more than two arms cannot be fully incorporated in the meta-analysis unless they are split into multiple two-arm trials. Those two-arm trials are treated as if they were independent; whereas they came from the same trial.* Consequently, such a network meta-analysis often incurs bias and loss of efficiency, as observed in Jansen et al. (2011) and Hoaglin et al. (2011). Taking into account this drawback, we consider $\widehat{\delta}_{AB,direct}$ and $\widehat{\delta}_{AB,indirect}$ as two separate estimators of δ_{AB} in the analysis in later sections.

We show later that the CD approach can combine the direct and indirect evidence for δ_{AB} efficiently, provided that the observed evidences from the direct and indirect comparisons are consistent with each other or at least not contradictory.

3.1.4 Bayesian hierarchical model

Similar to the CD approach, a Bayesian approach can also incorporate all trials. However, the Bayesian approach has to rely on prior distributions, which then impose additional assumptions.

To carry out network meta-analysis on clinical trials with direct and indirect treatment comparisons, Lu and Ades (2004, 2006) proposed the following hierarchical Bayesian model:

$$\begin{array}{ll} \text{level 1:} & r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), i = 1, 2, \ldots, 36, j = A, B, C \\ \text{level 2:} & (\delta_{AB,i}, \delta_{AC,i})^{\mathbf{T}}|\boldsymbol{\delta}, \mathbf{C} \sim N(\boldsymbol{\delta}, \mathbf{C}) \quad \perp \quad \mu_i|\mu, \sigma_\mu^2 \sim N(\mu, \sigma_\mu^2) \\ \text{level 3:} & \text{hyper prior distributions for } \boldsymbol{\delta}, \mathbf{C} \\ & \text{and parameters in the distribution of } \mu, \sigma_\mu^2 \text{ if necessary} \end{array} \tag{15}$$

where

$$\begin{bmatrix} \delta_{AB,i} \\ \delta_{AC,i} \\ \mu_i \end{bmatrix} = \mathbf{T}_{\texttt{BS}} \begin{bmatrix} \text{logit}(p_{iA}) \\ \text{logit}(p_{iB}) \\ \text{logit}(p_{iC}) \end{bmatrix} \text{ and } \mathbf{T}_{\texttt{BS}} \triangleq \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

As stated in Lu and Ades (2004), this model extends the one proposed by Smith et al. (1995) to address the issues of incorporating indirect comparisons and to fully incorporate trials with more than two arms.

Specifically, Lu and Ades (2004) considered two sets of prior distributions, Bayesian-HOM prior and Bayesian-HET prior. The first set of prior distributions ("Bayesian-HOM") assumes a homogenous variance for $\delta_{AB,i}$ and $\delta_{AC,i}$:

$$\begin{gathered} \boldsymbol{\delta} \sim N(0, 10^3 \mathbf{I}_2) \\ \mathbf{C} = \sigma^2 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}, \quad \sigma^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3}) \\ \mu \sim N(0, 10^3), \quad \sigma_\mu^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3}) \end{gathered} \tag{16}$$

The second set of prior distributions ("Bayesian-HET") allows heterogenous variances for $\delta_{AB,i}$ and $\delta_{AC,i}$:

$$\begin{gathered} \boldsymbol{\delta} \sim N(0, 10^3 \mathbf{I}_2) \\ \mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad \text{where } \rho = 0.5 \\ \sigma_j^2 \sim \text{Gamma}(a, b), a \sim \text{Exp}(0.01), b \sim \text{Gamma}(10^{-3}, 10^{-3}), j = 1, 2 \\ \mu \sim N(0, 10^3), \quad \sigma_\mu^{-2} \sim \text{Gamma}(10^{-3}, 10^{-3}) \end{gathered} \tag{17}$$

Except for the different assumptions on the structure of covariance matrix $\mathbf{C}$, both Bayesian-HOM and Bayesian-HET impose the same noninformative priors on $\boldsymbol{\delta}, \mu$, and σ_μ^2. The assumptions of priors are subjective and often difficult to verify. Our numerical studies in Section 4 suggest that the Bayesian approach is sensitive to the choice of priors.

3.1.5 Results

We consider the following six methods and compare their inferences on δ_{AB}:

- Traditional-Direct: Traditional frequentist meta-analysis on direct pairwise comparisons.
- Traditional-Indirect: Traditional frequentist meta-analysis on indirect pairwise comparisons.
- Bayesian-HOM: Bayesian network meta-analysis with homogeneous variance structure on $\boldsymbol{\delta}$.
- Bayesian-HET: Bayesian network meta-analysis with heterogeneous variance structure on $\boldsymbol{\delta}$.
- CD[$\mathbf{S}_{\texttt{DL}}$]: The proposed CD procedure with $\mathbf{S}$ estimated by an extension of the DerSimonian and Laird method to the multivariate case (Jackson et al. (2010)).
- CD[$\mathbf{S}_{\texttt{REML}}$]: The proposed CD procedure with $\mathbf{S}$ estimated by maximizing restricted likelihood.

The values of $\widehat{\delta}_{AB}$ and its corresponding 95% confidence interval (CI) or 95% credible interval (CrI) from all six methods are summarized in Table 2.

Table 2: Results of meta-analyses on CAD data

Method	$\widehat{\delta}_{AB}$	s.d.$(\widehat{\delta}_{AB})$	95% CI	Length of 95% CI
Traditional-Direct	-1.3757	0.1672	$(-1.7035, -1.0479)$	0.6556
Traditional-Indirect	-1.2874	0.5129	$(-2.2926, -0.2822)$	2.0104
Bayesian-HOM	-1.3681	0.1084	$(-1.5900, -1.1650)$	0.4250
Bayesian-HET	-1.3770	0.1312	$(-1.6170, -1.1028)$	0.5142
CD[$\mathbf{S}_{\texttt{DL}}$]	-1.2984	0.1174	$(-1.5285, -1.0683)$	0.4602
CD[$\mathbf{S}_{\texttt{REML}}$]	-1.2957	0.1096	$(-1.5104, -1.0809)$	0.4295

Table 2 shows that all six methods yield similar point estimates of δ_{AB}. However, because they use both direct and indirect evidence, the Bayesian methods and the CD methods yield smaller variance estimates and tighter confidence interval, in comparison with traditional pairwise meta-analysis. Also, the results from indirect comparisons are in line with those obtained from direct comparisons, although less efficient. It seems appropriate to combine the trials with direct and indirect evidence.

3.2 An example on cirrhosis

As another example, we consider the data presented in Pagliaro et al. (1992) and used in Lu and Ades (2004). The authors analyzed 26 trials of non-surgical treatments intended to prevent

first bleeding in patients with cirrhosis and esophageal varices who had never bled, in order to assess the effectiveness of three types of treatments: beta-blockers, endoscopic sclerotherapy and non-active treatment (control), denoted by A, B, and C, respectively. Of the 26 trials, 2 trials compared all three treatments, 7 trials compared beta-blockers vs control, and 17 trials compared sclerotherapy vs control. In Table 3, for trial i and treatment j, r_{ij} is the number of patients who had a first bleeding event and n_{ij} is the total number of patients. Our concern is with the relative performance of the active treatments: beta-blockers vs sclerotherapy. However, the only trials that compared them directly were the two three-arm trials, which were not sufficiently large. In this situation direct evidence is not strong enough, and incorporating indirect evidence is particularly important for making inferences.

Table 3: Cirrhosis data: number of patients who had a first bleeding event.

Study	Beta-blockers (A)		Sclerotherapy (B)		Control (C)	
	r_{ij}	n_{ij}	r_{ij}	n_{ij}	r_{ij}	n_{ij}
1	2	43	9	42	13	41
2	12	68	13	73	13	72
3	4	20	—	—	4	16
4	20	116	—	—	30	111
5	1	30	—	—	11	49
6	7	53	—	—	10	53
7	18	85	—	—	31	89
8	2	51	—	—	11	51
9	8	23	—	—	2	25
10	—	—	4	18	0	19
11	—	—	3	35	22	36
12	—	—	5	56	30	53
13	—	—	5	16	6	18
14	—	—	3	23	9	22
15	—	—	11	49	31	46
16	—	—	19	53	9	60
17	—	—	17	53	26	60
18	—	—	10	71	29	69
19	—	—	12	41	14	41
20	—	—	0	21	3	20
21	—	—	13	33	14	35
22	—	—	31	143	23	138
23	—	—	20	55	19	51
24	—	—	3	13	12	16
25	—	—	3	21	5	28
26	—	—	6	22	2	24

We apply the same six methods as in the CAD data set. The parameter of interest is δ_{AB}, the log-odds ratio of first bleeding for beta-blockers vs sclerotherapy. The results are presented in Table 4.

Table 4: Results of meta-analysis on cirrhosis data

Method	$\widehat{\delta}_{AB}$	s.d.$(\widehat{\delta}_{AB})$	95% CI	Length of 95% CI
Traditional-Direct	0.7284	0.8439	$(-0.9256, 2.3824)$	3.3080
Traditional-Indirect	-0.0927	0.8069	$(-1.6738, 1.4884)$	3.1622
Bayesian-HOM	0.5228	0.3171	$(-0.0969, 1.1461)$	1.2430
Bayesian-HET	0.6466	0.3250	(0.0410, 1.3151)	1.2741
CD[$\mathbf{S}_{\mathtt{DL}}$]	0.5688	0.2588	(0.0617, 1.0761)	1.0144
CD[$\mathbf{S}_{\mathtt{REML}}$]	0.6381	0.2445	(0.1589, 1.1174)	0.9585

In Table 4, we again observe that the Bayesian methods and the CD procedures have substantially lower variance as a result of integrating all treatment comparisons. Therefore, the network-meta-analysis approaches have effectively strengthened the results obtained from direct comparisons by borrowing information from indirect comparisons. Unlike the results in the CAD example, pairwise meta-analysis using only direct comparisons does not achieve significant results, whereas the Bayesian and CD approaches yield significant or almost significant results. However, the validity of combining direct and indirect treatment comparisons should be carefully investigated, the difference between $\widehat{\delta}_{AB,indirect}$ and $\widehat{\delta}_{AB,direct}$ raises concerns about consistency between direct and indirect evidence. The topic of inconsistent evidence is discussed in Higgins et al. (2002, 2003). We also discuss this topic further in Section 4.3 and Section 5.

In these two examples, the CD and Bayesian approaches yield similar results. The confidence intervals derived from the CD approach are only slightly tighter than those derived from the Bayesian approach. However, our simulation studies in the next section show that the Bayesian credible intervals may not achieve the nominal coverage probability, and their empirical coverage probabilities may be far below the nominal level when the assumed prior on the between-trial covariance structure does not agree with the underlying true model. This latter condition is almost impossible to verify in practice. In contrast, the proposed CD combining approach does not require any prior, and the derived confidence intervals can maintain adequate coverage probability regardless of the between-trial covariance structure.

4 Simulation studies

We conducted simulation studies to compare the performance of the proposed CD combining approach with traditional pairwise meta-analysis and the Bayesian method.

4.1 Simulation settings

We based our simulation on the structure of the cirrhosis data. Specifically, the evidence network involves three treatments (A, B, and C). The problem of interest is to infer the

relative effectiveness of A vs B.

Consider two scenarios, one with 24 trials and the other with 96 trials. In the first scenario, the 24 clinical trials, comprise 1 trial comparing all three treatments, 3 trials comparing A and B, 10 trials comparing treatments A and C, and 10 trials comparing B and C. The number of patients in each arm of each trial is 100, i.e., $n_{ij} = 100, \forall i$ and $j \in T_i$. In the second scenario the number of trials of each type is four times that in the first scenario. The simulation is designed to show the benefit of borrowing strength from indirect evidence when direct evidence (trials directly comparing treatments A and B) is somewhat limited.

Table 5: Simulation Settings - Number of Trials k and Patients Involved in Each Group n_{ij}

Total Number of Trials k \Type of Trial		ABC	AB	AC	BC	n_{ij}
Simulation Scenario 1	k=24	1	3	10	10	100
Simulation Scenario 2	k=96	4	12	40	40	100

We generate the simulated data from the model:

$$\begin{gathered} r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad p_{ij} = \frac{\exp(\theta_{ij})}{1+\exp(\theta_{ij})}, \quad i = 1, 2, \ldots, 24 \text{ or } 96, j \in T_i \\ \boldsymbol{\theta}_i \sim N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}) \end{gathered} \tag{18}$$

where $\mathbf{A}_i$ consists of the rows of the identity matrix corresponding to the treatments in T_i.

We specify the true value of $\boldsymbol{\theta} = (-1.82, -1.21, -0.80)^{\mathbf{T}}$ as the values are close to those estimated from the cirrhosis data. It follows that the probabilities of observing an event in treatment A, B, and C are $\mathbf{p} = (0.14, 0.23, 0.31)^{\mathbf{T}}$. For the covariance matrix $\mathbf{S}$, we consider three cases:

Case 1:

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \iff \mathbf{B} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix};$$

Case 2:

$$\mathbf{S} = \begin{bmatrix} 2.5736 & -1.2868 & 1.7132 \\ -1.2868 & 4.8528 & -0.5660 \\ 1.7132 & -0.5660 & 1.8528 \end{bmatrix} \iff \mathbf{B} = \begin{bmatrix} 10 & 1.5811 & 0 \\ 1.5811 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

Case 3:

$$\mathbf{S} = \begin{bmatrix} 3.1070 & 0.4314 & 1.2358 \\ 0.4314 & 0.7557 & 0.4693 \\ 1.2358 & 0.4693 & 0.8645 \end{bmatrix} \iff \boldsymbol{B} = \begin{bmatrix} 3.0000 & 1.9092 & -1.0392 \\ 1.9092 & 1.5000 & -0.7348 \\ -1.0392 & -0.7348 & 1.0000 \end{bmatrix},$$

where $\mathbf{B} = \text{cov}(\delta_{AB,i}, \delta_{AC,i}, \mu_i) = \mathbf{T}_{\mathtt{BS}}\mathbf{S}\mathbf{T}_{\mathtt{BS}}^{\mathbf{T}}$, and $\delta_{AB,i}, \delta_{AC,i}, \mu_i$ and $\mathbf{T}_{\mathtt{BS}}$ are defined as in model (15). Here " $\iff$ " indicates the one-to-one correspondence between the covariance matrix $\mathbf{S}$ in model (18) and the covariance matrix $\mathbf{B}$ in the Bayesian models.

In Case 1, $\mathbf{S}$ is set to an identity matrix to ensure that the true model (18) meets the assumptions of Bayesian-HOM in Section 3.1.4, and is thus equivalent to the case of (16) with $\sigma^2 = 2$.

Similarly, the covariance matrix $\mathbf{S}$ in Case 2 allows the true model (18) to meet the assumptions of Bayesian-HET, and is thus equivalent to the case of $\sigma_1^2 = 10, \sigma_2^2 = 1$ and $\rho = 0.5$ in (17). As suggested in Joseph et al. (1997), we further extend the model to incorporate correlations between $\delta_{AB,i}, \delta_{AC,i}$ and μ_i, instead of assuming independence. Therefore, in Case 3, the covariance matrix $\mathbf{S}$ is specified to give an arbitrary covariance structure such that $\mathbf{B}$ fails to meet the assumptions of either Bayesian-HOM or Bayesian-HET.

In summary, we consider a total of six ($= 2 \times 3$) settings in our simulation study: 24 and 96 trials each with three specifications of the covariance matrix $\mathbf{S}$.

4.2 Results

We consider and compare the performance of a total of nine approaches. They include the six methods listed in Section 3.1.5: Traditional-Direct and Traditional-Indirect, Bayesian-HOM and Bayesian-HET, and CD[$\mathbf{S}_{\texttt{DL}}$] and CD[$\mathbf{S}_{\texttt{REML}}$]. Additionally, we include three other CD approaches: two semi-Bayesian approaches, CD[$\mathbf{S}_{\texttt{BHOM}}$] and CD[$\mathbf{S}_{\texttt{BHET}}$], in which the covariance matrix $\mathbf{S}$ is estimated by the Bayesian method with prior in (16) and (17), respectively, and CD[$\mathbf{S}_{\texttt{TRUE}}$], which uses the true covariance matrix $\mathbf{S}$. The CD[$\mathbf{S}_{\texttt{TRUE}}$] method allows us to separate the effect of estimating the mean alone and study the potential impacts on estimation of the mean when different approaches are used to estimate $\mathbf{S}$. Thus, the nine methods are:

- Traditional frequentist methods:
  - Traditional-Direct: Traditional frequentist meta-analysis of direct pairwise comparisons.
  - Traditional-Indirect: Traditional frequentist meta-analysis via indirect pairwise comparisons.
- Bayesian methods:
  - Bayesian-HOM: Bayesian network meta-analysis with homogenous variance structure on $\boldsymbol{\delta}$.
  - Bayesian-HET: Bayesian network meta-analysis with heterogenous variance structure on $\boldsymbol{\delta}$.
- CD methods:
  - CD[$\mathbf{S}_{\texttt{DL}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{DL}}$.
  - CD[$\mathbf{S}_{\texttt{REML}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{REML}}$.
  - CD[$\mathbf{S}_{\texttt{BHOM}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{BHOM}}$.
  - CD[$\mathbf{S}_{\texttt{BHET}}$]: $\mathbf{S}$ estimated by $\mathbf{S}_{\texttt{BHET}}$.
  - CD[$\mathbf{S}_{\texttt{TRUE}}$]: using the known true $\mathbf{S}$.

In simulation Scenario 1 Case 1, for example, we generate data according to the model specified in (18), and then apply each method to estimate δ_{AB} and calculate the corresponding 95% confidence (credible) interval. We repeat this process 1000 times. For each method, we report the mean and standard deviation of the 1000 $\widehat{\delta}_{AB}$ and the percentage of times (coverage) that

the 1000 95% CIs cover the true $\delta_{AB} = 0.6070$ and the average interval length. The results for Scenarios 1 and 2 with Case 1 ($\mathbf{S} = \mathbf{I}_{3\times 3}$) are presented in Table 6. Similarly, the results for Case 2 and Case 3 are presented in Tables 7 and 8. It is straightforward to verify that the chance that no trial has zero events in the entire 1000 replications is at least 99.97%. Thus the zero events issue is not considered in the simulation study.

Table 6: Summary of results of simulation studies - Case 1

Method	$\widehat{\delta}_{AB}$	s.d.$(\widehat{\delta}_{AB})$	95% CI coverage	Average Length of 95% CI
Scenario 1 - Small Number of Trials $k = 24$				
Traditional-Direct	0.5952	0.7167	0.867	2.7041
Traditional-Indirect	0.5913	0.6312	0.941	2.5225
Bayesian-HOM	0.5796	0.4097	0.937	1.5704
Bayesian-HET	0.5736	0.4104	0.938	1.5712
CD[$\mathbf{S}_{\texttt{REML}}$]	0.5677	0.4057	0.897	1.3766
CD[$\mathbf{S}_{\texttt{TRUE}}$]	0.5732	0.3850	0.955	1.5554
CD[$\mathbf{S}_{\texttt{DL}}$]	0.5718	0.4195	0.862	1.2550
CD[$\mathbf{S}_{\texttt{BHOM}}$]	0.5719	0.3925	0.940	1.5337
CD[$\mathbf{S}_{\texttt{BHET}}$]	0.5714	0.3927	0.943	1.5225
Scenario 2 - Large Number of Trials $k = 96$				
Traditional-Direct	0.5843	0.3658	0.927	1.3950
Traditional-Indirect	0.6104	0.3118	0.962	1.2681
Bayesian-HOM	0.6126	0.2016	0.948	0.7663
Bayesian-HET	0.6126	0.2016	0.943	0.7701
CD[$\mathbf{S}_{\texttt{REML}}$]	0.5780	0.1915	0.936	0.7242
CD[$\mathbf{S}_{\texttt{TRUE}}$]	0.5856	0.1900	0.966	0.7777
CD[$\mathbf{S}_{\texttt{DL}}$]	0.5762	0.1932	0.904	0.6536
CD[$\mathbf{S}_{\texttt{BHOM}}$]	0.5852	0.1920	0.959	0.7716
CD[$\mathbf{S}_{\texttt{BHET}}$]	0.5852	0.1918	0.954	0.7680

From the results in Tables 6, 7 and 8, it is evident that the traditional pairwise meta-analysis is much less efficient than the CD network meta-analysis approaches. Specifically, compared with the results from the CD[$\mathbf{S}_{\texttt{REML}}$] method, the lengths of 95% CIs obtained from traditional meta-analysis methods are much greater, even though the probabilities of covering the true value are comparable. This suggests that, when the parameter of interest is a vector, information on one parameter may be potentially useful for inferences on other parameters. Thus, mixed treatment comparisons should be considered in our settings.

Consider the probability that the nominal 95% CI covers the true δ_{AB} as one criterion for assessing the performance of each meta-analysis method. It is evident from the simulation study that the results of the Bayesian methods are sensitive to the specifications of their prior distributions. Specifically, Bayesian-HOM fails to achieve appropriate coverage in Cases 2 and 3 (e.g., 89% and 90% in Table 7 and 86% and 81% in Table 8), regardless whether the number of studies is small or large. Similarly, Bayesian-HET fails to provide satisfactory coverage in

Table 7: Summary of results of simulation studies - Case 2

Method	$\widehat{\delta}_{AB}$	s.d.($\widehat{\delta}_{AB}$)	95% CI coverage	Average Length of 95% CI
Scenario 1 - Small Number of Trials $k = 24$				
Traditional-Direct	0.6176	1.4759	0.849	5.4220
Traditional-Indirect	0.5905	0.8818	0.937	3.4753
Bayesian-HOM	0.6095	0.7450	0.887	2.4177
Bayesian-HET	0.5706	0.7360	0.913	2.6355
CD[$\mathbf{S}_{\texttt{REML}}$]	0.5793	0.6922	0.916	2.5426
CD[$\mathbf{S}_{\texttt{TRUE}}$]	0.5820	0.6865	0.973	2.9649
CD[$\mathbf{S}_{\texttt{DL}}$]	0.6165	0.7289	0.811	2.0011
CD[$\mathbf{S}_{\texttt{BHOM}}$]	0.6323	0.7030	0.901	2.3815
CD[$\mathbf{S}_{\texttt{BHET}}$]	0.6044	0.6930	0.906	2.4856
Scenario 2 - Large Number of Trials $k = 96$				
Traditional-Direct	0.6433	0.7431	0.924	2.8474
Traditional-Indirect	0.6287	0.4279	0.951	1.7643
Bayesian-HOM	0.6852	0.3540	0.899	1.1858
Bayesian-HET	0.6454	0.3436	0.960	1.3952
CD[$\mathbf{S}_{\texttt{REML}}$]	0.6200	0.3226	0.959	1.3164
CD[$\mathbf{S}_{\texttt{TRUE}}$]	0.6261	0.3254	0.980	1.4823
CD[$\mathbf{S}_{\texttt{DL}}$]	0.6455	0.3227	0.864	0.9721
CD[$\mathbf{S}_{\texttt{BHOM}}$]	0.6636	0.3324	0.933	1.2085
CD[$\mathbf{S}_{\texttt{BHET}}$]	0.6279	0.3256	0.968	1.3876

the Case 3 (85% and 82% in Table 8) when its assumption on prior cannot cover the true model. In summary, both Bayesian methods are able to estimate δ_{AB} properly only if their prior assumptions cover the underlying true covariance model, and they fail to do so when their prior assumptions are not compatible with the underlying true covariance model. So the Bayesian procedures are vulnerable to their assumptions on priors, and we should make as few assumptions as possible when specifying priors.

In examining the results of the CD procedures, we first observe that CD[$\mathbf{S}_{\texttt{TRUE}}$] achieves desirable coverage rates in all cases (95%−98% in Tables 6, 7, and 8). Therefore, the performance of the CD procedure is satisfactory for combining information on $\boldsymbol{\theta}$. However, the performance of the CD procedure is strongly affected by the quality of estimating the covariance matrix $\mathbf{S}$. To help establish a practical guideline, we compare the quality of estimates based on the extended DL method $\mathbf{S}_{\texttt{DL}}$ and the REML method $\mathbf{S}_{\texttt{REML}}$. Specifically, we plug in the corresponding estimates in the process of constructing and combining individual CDs, and again we study the performance of estimates $\widehat{\delta}_{AB}$ and the corresponding 95% CIs. The performance of CD[$\mathbf{S}_{\texttt{REML}}$] is reasonable in all settings, i.e., close to the nominal 95% coverage (see, e.g., 92% − 96% in Tables 6, 7, and 8) as long as the number of studies is sufficiently large. Further, the coverage rate of CD[$\mathbf{S}_{\texttt{REML}}$] improves from 89% − 92% to 92% − 96% as the number of studies increases from 24 to 96. On the other hand, the coverage rate of CD[$\mathbf{S}_{\texttt{DL}}$] is relatively low, around

Table 8: Summary of results of simulation studies - Case 3

Method	$\widehat{\delta}_{AB}$	s.d.$(\widehat{\delta}_{AB})$	95% CI coverage	Average Length of 95% CI
Scenario 1 - Small Number of Trials $k = 24$				
Traditional-Direct	0.4706	0.8260	0.868	3.0721
Traditional-Indirect	0.4250	0.4582	0.915	1.8193
Bayesian-HOM	0.4135	0.4400	0.855	1.4116
Bayesian-HET	0.4065	0.4388	0.853	1.4186
CD[$\mathbf{S}_{\texttt{REML}}$]	0.4834	0.4201	0.892	1.4924
CD[$\mathbf{S}_{\texttt{TRUE}}$]	0.5010	0.4058	0.953	1.7241
CD[$\mathbf{S}_{\texttt{DL}}$]	0.3957	0.4510	0.787	1.2756
CD[$\mathbf{S}_{\texttt{BHOM}}$]	0.3750	0.4169	0.855	1.3811
CD[$\mathbf{S}_{\texttt{BHET}}$]	0.3753	0.4141	0.852	1.3824
Scenario 2 - Large Number of Trials $k = 96$				
Traditional-Direct	0.4823	0.4132	0.912	1.5936
Traditional-Indirect	0.4472	0.2250	0.896	0.9051
Bayesian-HOM	0.4603	0.2131	0.807	0.6828
Bayesian-HET	0.4589	0.2097	0.822	0.6996
CD[$\mathbf{S}_{\texttt{REML}}$]	0.5057	0.1943	0.919	0.7724
CD[$\mathbf{S}_{\texttt{TRUE}}$]	0.5261	0.1978	0.949	0.8620
CD[$\mathbf{S}_{\texttt{DL}}$]	0.4435	0.2029	0.749	0.6242
CD[$\mathbf{S}_{\texttt{BHOM}}$]	0.3959	0.2027	0.754	0.6954
CD[$\mathbf{S}_{\texttt{BHET}}$]	0.3950	0.2002	0.759	0.7042

79% − 86%, when the sample size is small. Moreover, the performance of CD[$\mathbf{S}_{\texttt{DL}}$] does not always improve as the number of studies increases. For example, the coverage rate of CD[$\mathbf{S}_{\texttt{DL}}$] drops from 78.7% to 74.9% in Table 8. Thus, the REML method is preferable to the extended DL method for estimating the covariance matrix $\mathbf{S}$. This observation is consistent with the shortcomings of the DL method reported in univariate random-effects models by Emerson et al. (1993). Between the REML and DL methods, we recommend the CD procedure with $\mathbf{S}_{\texttt{REML}}$ for network meta-analysis when $\mathbf{S}$ is unknown.

Finally, the results for the semi-Bayesian CD procedures appear to be similar to the results for the corresponding Bayesian procedures. Specifically, the performance of CD[$\mathbf{S}_{\texttt{BHOM}}$] is in line with Bayesian-HOM. It achieves appropriate coverage in Case 1 (94% and 96% in Table 6), but fails in Cases 2 and 3 (90% and 93% in Table 7 and 86% and 75% in Table 8), regardless of the number of studies $k = 24$ or 96. Similarly, the results for CD[$\mathbf{S}_{\texttt{BHET}}$] are in line with Bayesian-HET. It provides satisfactory coverage in Cases 1 and 2 (94% and 95% in Table 6 and 91% and 97% in Table 7), but fails Case 3 (85% and 76% in Table 8). Once again, the CD procedure is sensitive to the quality of estimation of $\mathbf{S}$. Also, the confidence distribution $H^{(c)}(\cdot)$ in (8)is an asymptotic CD that is more suitable for making inferences on $\boldsymbol{\theta}$ when $k \to \infty$, under which both the mean vector $\boldsymbol{\theta}$ and the between-trials covariance matrix $\mathbf{S}$ can be estimated consistently.

4.3 A CD approach with adaptive weights

As we observed from in Section 4.2, the overall findings for a network can be quite unreliable when indirect evidence and direct evidence inconsistent. In this section, an adaptive weighting system improves resistance to the impact of inconsistent indirect comparisons by down-weighting the trials that contribute to the inconsistent evidence. Here, the degree of inconsistency from an indirect comparison is measured by how the trials in the indirect comparison deviate from the overall outcome for the direct comparison. The precise formulation of this measure, which we loosely call "distance," is given after Model (19). Taking into account this distance, the CD combining process can still use indirect comparisons that provide outcomes consistent with those from the direct comparisons, but it can also reduce the impact of inconsistent indirect comparisons. We demonstrate this property through the following simulation studies.

We consider the model (18) used in Scenario 1 in Section 4.1, with two modifications. First, we increase the total number of trials from 24 to 33 so that three trials, instead of one trial, compare treatments A, B, and C, and ten trials, instead of three trials, directly compare treatments A and B. We still have ten trials comparing treatments A and C and ten trials comparing treatments B and C. Thus, for inferences on δ_{AB}, we have 13 direct comparisons and 20 trials with information on the indirect comparison. Second, the trials containing information on the direct comparison are consistent, but some of the remaining 20 trials containing information on the indirect comparison may be biased. Specifically, we consider the following model to generate the simulation data:

$$
\begin{aligned}
r_{ij}|p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}), \quad p_{ij} = \frac{\exp(\theta_{ij})}{1+\exp(\theta_{ij})}, \quad i = 1, 2, \ldots, 33, \quad j \in T_i \\
\boldsymbol{\theta}_i \sim (1-\epsilon) N(\mathbf{A}_i\boldsymbol{\theta}, \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}) + \epsilon N(\mathbf{A}_i(\boldsymbol{\theta}-\boldsymbol{\eta}_i), \mathbf{A}_i\mathbf{S}\mathbf{A}_i^{\mathbf{T}}) \\
\text{where} \\
\epsilon = 0 \text{ and } \boldsymbol{\eta}_i = \mathbf{0} \text{ for } i \text{ s.t. } T_i = \{A, B, C\} \text{ or } T_i = \{A, B\} \\
\epsilon = 0.4 \text{ and } \boldsymbol{\eta}_i = \begin{cases} (\eta_{A,i}, 0, 0)^{\mathbf{T}} & \text{for } i \text{ s.t. } T_i = \{A, C\} \\ (0, \eta_{B,i}, 0)^{\mathbf{T}} & \text{for } i \text{ s.t. } T_i = \{B, C\} \end{cases}
\end{aligned}
\tag{19}
$$

Here, the values of $\eta_{A,i}$ and $\eta_{B,i}$ are fixed numbers simulated from $N(2, 4)$.

Model (19) indicates that all trials that compare both treatments A and B directly have the same underlying true parameter $\boldsymbol{\theta}$, whereas some trials involving A only or B only may have different underlying true parameters. If we are to include the trials that provide the indirect comparison in our analysis, it would be desirable to exclude or down-weight those trials. In this case, we devise the following notion of distance d_i,

$$
d_i = \begin{cases} \dfrac{(\widehat{\delta}_{AC,i} - \text{median}_{l \, s.t. \, T_l = \{B,C\}} \widehat{\delta}_{BC,l}) - \widehat{\delta}_{AB,direct}}{\sqrt{\text{var}(\widehat{\delta}_{AB,direct})}} & \text{for } i \text{ s.t. } T_i = \{A, C\} \\ \dfrac{(\text{median}_{l \, s.t. \, T_l = \{A,C\}} \widehat{\delta}_{AC,l} - \widehat{\delta}_{BC,i}) - \widehat{\delta}_{AB,direct}}{\sqrt{\text{var}(\widehat{\delta}_{AB,direct})}} & \text{for } i \text{ s.t. } T_i = \{B, C\}, \end{cases}
$$

where $\widehat{\delta}_{AB,direct}$ and $\text{var}(\widehat{\delta}_{AB,direct})$ are obtained from Equation (13). Heuristically, d_i for each indirect comparison trial measures its deviation from the overall outcome given by all direct

comparison trials. For example, we could consider including only the studies with distance $|d_i| \leq 1$ in the meta-analysis. In other words, we would define w_i^* as

$$w_i^* = \begin{cases} 1 & \text{if } |d_i| \leq 1 \\ 0 & \text{if } |d_i| > 1, \end{cases}$$

and use w_i^* in the method CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted. Specifically, we set $W_i = w_i^* \times \mathbf{A}_i^+(\widehat{\mathbf{\Sigma}}_i + \mathbf{A}_i\mathbf{S}_{\texttt{REML}}\mathbf{A}_i^{\mathbf{T}})^{-1}\mathbf{A}_i$, and take the cdf of the random vector in (7) as the combined multivariate normal CD. We show that in this way the combined CD is able to exclude those inconsistent indirect trials – trials with large d_i. There are many other choices of adaptive weights. For convenience, we use here the simple, though somewhat restrictive, $|d_i| \leq 1$ to remove inconsistent studies from combination. A detailed discussion of choices of adaptive weights and their applications to combining CDs can be found in Xie et al. (2011).

In a further simulation study (Case 4), we consider two settings. In Setting 1, we generate the simulated data using model (18), in which all studies have the same underlying true parameter value, but modify it to have 33 trials with the same composition of trials as model (19). In Setting 2, the simulated data are generated from model (19). In this case, some trials used in the indirect comparison have a different underlying true parameter value. In both settings, three trials compare all three treatments, ten trials compare treatments A and B, ten trials A and C, and ten trials B and C. The number of patients involved in each arm of each study is 100. We apply CD[$\mathbf{S}_{\texttt{REML}}$], CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted, and CD[$\mathbf{S}_{\texttt{TRUE}}$] to the simulated data sets. We repeat the entire process 1000 times and report the results in Table 9.

Table 9: Summary of results of simulation studies - Case 4

Method	$\widehat{\delta}_{AB}$	s.d.($\widehat{\delta}_{AB}$)	95% CI coverage	Average Length of 95% CI
Setting 1 - 33 Trials without Inconsistent Indirect Trials				
CD[$\mathbf{S}_{\texttt{REML}}$]	0.5733	0.2984	0.9200	1.1122
CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted	0.5780	0.3705	0.9230	1.4078
CD[$\mathbf{S}_{\texttt{TRUE}}$]	0.5818	0.2955	0.9520	1.2139
Setting 2 - 33Trials with Inconsistent Indirect Trials				
CD[$\mathbf{S}_{\texttt{REML}}$]	1.1425	0.3932	0.7190	1.4808
CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted	0.6479	0.3934	0.9770	1.9963
CD[$\mathbf{S}_{\texttt{TRUE}}$]	1.1001	0.3367	0.6260	1.2250

All three methods are able to achieve appropriate coverage rate ($92\% - 95\%$ in Setting 1) if all trial outcomes are consistent with one another. However, in Setting 2, with inconsistent indirect trials, only CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted provides appropriate inference on δ_{AB}. In particular, the estimate $\widehat{\delta}_{AB} = 0.6479$ by CD[$\mathbf{S}_{\texttt{REML}}$]-adjusted is not far from the true $\delta_{AB} = 0.6070$, and its 95% CI has a coverage rate of 97.7%. Therefore, with carefully designed study-specific weights, the CD procedure is able to provide some resistance to the impact of inconsistent indirect trials mistakenly included in the meta-analysis.

5 Concluding remarks

In this paper, we have proposed a frequentist method for network meta-analysis by combining multivariate normal confidence distributions (CDs) associated with individual studies. This proposed CD approach can perform indirect comparisons in a network of mixed treatment comparisons, and it can use the findings from indirect comparisons efficiently to enhance the overall inference of the entire network. The CD approach can also be modified by using an adaptive weighting scheme to reduce the effect of indirect comparisons whose findings contradict those from the direct comparisons. Overall, the proposed CD approach can effectively and efficiently integrate direct and indirect information from disparate sources. In fact, the CD approach can estimate consistently and efficiently the parameters of interest as well as the between-trials covariance matrix when the number of studies goes to infinity. Through simulation studies, we have also demonstrated that the CD approach generally outperforms traditional pairwise meta-analysis and the Bayesian hierarchical model. In conclusion, the CD approach is highly competitive for network meta-analysis.

In comparing the approaches on the CAD data in Section 3.1, we excluded the TAXUS I trial to avoid addressing the issue of zero events there. In traditional pairwise meta-analysis, one customarily adds 0.5 to zero events. This correction is arbitrary and introduces bias in the inferences. By removing zero-event trials from the analysis, one would lose the information they contain. For example, for TAXUS I, zero event is a favorable outcome for both BMS and PES. This loss can cause concerns as well, especially if the zero-event trials constitute a sizable portion of the data. For an exact inference method involving zero events, the approach of combining significance functions proposed in Liu et al. (2012) can avoid the shortcomings of the earlier approaches.

In network meta-analysis, it is important to assess the consistency of the evidence from all trials in the network. However, such assessment is often difficult. One reason is that designs often differ between the trials yielding direct comparisons and the trials leading to indirect comparisons. Furthermore, it is practically impossible to distinguish between inconsistency and heterogeneity of random effects. See Higgins et al. (2002, 2003) for further discussion of this topic.

Although our examples involve clinical trials in medical studies, we emphasize that the proposed CD approach can be applied broadly for any multiple comparison studies in many other domains. For example, to establish ratings for a list of restaurants based on a survey of customer ratings, customers would be able to provide data only on the restaurants that they have patronized. The CD approach could be applied by constructing and combining CDs based on the ratings given to those restaurants by a group of customers.

6 Appendix

Lemma 1 Suppose $W_i, i = 1, \ldots, k$ are $p \times p$ positive semi-definite symmetric matrices and V_i is the column space of W_i. Let $V = V_1 + V_2 + \cdots + V_k \overset{\triangle}{=} \{\sum_{i=1}^{k} v_i | v_i \in V_i, i = 1, \ldots, k\}$. Then

$\sum_{i=1}^{k} W_i$ is positive definite provided that $V = \Re^p$.

Proof of Lemma 1:

It is a direct result that $\sum_{i=1}^{k} W_i$ is positive semi-definite. Suppose there exists a $p \times 1$ vector $\mathbf{v} \neq 0$ such that $\mathbf{v}^\mathbf{T}(\sum_{i=1}^{k} W_i)\mathbf{v} = 0$. Then, for any fixed i, we have $\mathbf{v}^\mathbf{T} W_i \mathbf{v} = 0$, which implies that $W_i^{1/2}\mathbf{v} = 0$. It follows that $\mathbf{v} \in \text{kernel}(W_i^{1/2})$, and immediately $\mathbf{v} \in \text{kernel}(W_i)$ since W_i is symmetric. Thus $\mathbf{v} \perp V_i$. Since i is arbitrary, we conclude that $\mathbf{v} \perp V = \Re^p$ and $\mathbf{v}$ has to be 0, which contradicts the assumption that $\mathbf{v} \neq 0$.

Proof of Theorem 1:

Let $\boldsymbol{\xi}^{(c)} = (\sum_{i=1}^{k} W_i)^{-1} \sum_{i=1}^{k} W_i \mathbf{A}_i^{+} \boldsymbol{\xi}_i$ and $H^{(c)}(\mathbf{t}) = \Pr\{\boldsymbol{\xi}^{(c)} \leq \mathbf{t} | \mathbf{Y}_1, \ldots, \mathbf{Y}_k\}$. We need to show that $H^{(c)}(\cdot) = H(\mathbf{Y}_1, \ldots, \mathbf{Y}_k; \cdot)$ is a multivariate normal CD for $\boldsymbol{\theta}$. Define $H_{\boldsymbol{\lambda}}(t) = \Pr\{\boldsymbol{\lambda}^\mathbf{T}\boldsymbol{\xi}^{(c)} \leq t | \mathbf{Y}_1, \ldots, \mathbf{Y}_k\}$ for any given vector $\boldsymbol{\lambda}$ satisfying $||\boldsymbol{\lambda}||_2 = 1$. By Definition 2, it suffices to show that $H_{\boldsymbol{\lambda}}(t)$ is a univariate normal CD function for $\boldsymbol{\lambda}^\mathbf{T}\boldsymbol{\theta}$.

To do so, we first note that $H_{\boldsymbol{\lambda}}(t)$ goes from 0 to 1 monotonically as t goes from $-\infty$ to ∞. Thus, $H_{\boldsymbol{\lambda}}(t)$ is a cdf. Second, we note that $\boldsymbol{\xi}_i$, defined by $\boldsymbol{\xi}_i | \boldsymbol{Y}_i = \boldsymbol{y}_i \sim N(\boldsymbol{y}_i, \text{var}(\boldsymbol{Y}_i))$, is a CD random vector for $\boldsymbol{\theta}_i$, and furthermore, $\mathbf{A}_i^{+}\boldsymbol{\xi}_i$ is a CD random vector for $\boldsymbol{\theta}$ in the sense that the distribution function of $\boldsymbol{\eta}^\mathbf{T}\mathbf{A}_i^{+}\boldsymbol{\xi}_i$ is a CD for $\boldsymbol{\eta}^\mathbf{T}\boldsymbol{\theta}$ for any $\boldsymbol{\eta} \in V_i$. Since $(\sum_{i=1}^{k} W_i)^{-1}$ exists by Lemma 1, we consider the conditional distribution of $\left(W_i(\sum_{i=1}^{k} W_i)^{-1}\boldsymbol{\lambda}\right)^\mathbf{T} \mathbf{A}_i^{+}\boldsymbol{\xi}_i$ given $\mathbf{Y}_i$. Clearly, it is a univariate normal CD for $\left(W_i(\sum_{i=1}^{k} W_i)^{-1}\boldsymbol{\lambda}\right)^\mathbf{T} \boldsymbol{\theta}$, because $W_i(\sum_{i=1}^{k} W_i)^{-1}\boldsymbol{\lambda} \in V_i$. Therefore, it is straightforward to show that, at the true parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\Pr\{H_{\boldsymbol{\lambda}}(\mathbf{Y}_1, \ldots, \mathbf{Y}_k) \leq s\} = \Pr\left\{ \Phi\left(\frac{\sum_{i=1}^{k} \left(W_i(\sum_{i=1}^{k} W_i)^{-1}\boldsymbol{\lambda}\right)^\mathbf{T} \mathbf{A}_i^{+}\mathbf{Y}_i - \boldsymbol{\lambda}^\mathbf{T}\boldsymbol{\theta}_0}{\sqrt{\sum_{i=1}^{k} \sigma_i^2}} \right) \leq s \right\} = s$$

where $\sigma_i^2 = \text{var}\left(\left(W_i(\sum_{i=1}^{k} W_i)^{-1}\boldsymbol{\lambda}\right)^\mathbf{T} \mathbf{A}_i^{+}\boldsymbol{\xi}_i\right)$. Thus, we have established that, at the true $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and as a function of the sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_k$, $H_{\boldsymbol{\lambda}}(\mathbf{Y}_1, \ldots, \mathbf{Y}_k)$ follows the uniform distribution $U[0, 1]$. This completes the proof.

References

Cox, D. (2013). Discussion of "Confidence distribution, the frequentist distribution estimator of a parameter — A Review" by Xie and Singh. *International Statistical Review. In press.*

De Blasi, P. and Schweder, T. (2012). Tail symmetry of confidence curves based on log-likelihood ratio. *Technical Report.*

Dempster, A. P. (2008). The Dempster-Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, 48:365–377.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3):177–188.

Dias, S., Welton, N. J., Caldwell, D. M., and Ades, A. E. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine*, 29(7-8):932–944.

Emerson, J. D., Hoaglin, D. C., and Mosteller, F. (1993). A modified random-effect procedure for combining risk difference in sets of 2×2 tables from clinical trials. *Journal of the Italian Statistical Society*, 2(3):269–290.

Hannig, J. and Xie, M. (2012). A note on Dempster-Shafer recombination of confidence distributions. *Electron. J. Statist*, 6:1943 – 1966.

Hardy, R. and Thompson, S. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, 15(6):619–629.

Higgins, J., Thompson, S., Deeks, J., and Altman, D. (2002). Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *Journal of Health Services Research & Policy*, 7(1):51.

Higgins, J., Thompson, S., Deeks, J., and Altman, D. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414):557–560.

Hoaglin, D. C., Hawkins, N., Jansen, J. P., Scott, D. A., Itzler, R., Cappelleri, J. C., Boersma, C., Thompson, D., Larholt, K. M., Diaz, M., and Barrett, A. (2011). Conducting indirect-treatment-comparison and network-meta-analysis studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 2. *Value in Health*, 14(4):429–437.

Jackson, D., White, I., and Thompson, S. (2010). Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine*, 29(12):1282–1297.

Jansen, J. P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., Lee, K., Boersma, C., Annemans, L., and Cappelleri, J. C. (2011). Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. *Value in Health*, 14(4):417–428.

Joseph, L., Du Berger, R., and Belisle, P. (1997). Bayesian and mixed bayesian/likelihood criteria for sample size determination. *Statistics in Medicine*, 16(7):769–781.

Liu, D., Liu, R., and Xie, M. (2012). An exact meta-analysis approach and its application to 2x2 tables with rare events. *Preprint*.

Lu, G. and Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, 23(20):3105–3124.

Lu, G. and Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 101(474):447–459.

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, 21(16):2313–2324.

Martin, R. and Liu, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc., In press.*

Normand, S. (1999). Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18(3):321–359.

Pagliaro, L., D'Amico, G., Sörensen, T. I. A., Lebrec, D., Burroughs, A. K., Morabito, A., Tiné, F., Politi, F., and Traina, M. (1992). Prevention of first bleeding in cirrhosis. a meta-analysis of randomized trials of nonsurgical treatment. *Annals of Internal Medicine*, 117(1):59–70.

Schweder, T. and Hjort, N. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics*, 29(2):309–332.

Sidik, K. and Jonkman, J. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine*, 26(9):1964–1981.

Simmonds, M. and Higgins, J. (2007). Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine*, 26(15):2982–2999.

Singh, K., Xie, M., and Strawderman, W. E. (2005). Combining information from independent sources through confidence distributions. *The Annals of Statistics*, 33(1):159–183.

Singh, K., Xie, M., and Strawderman, W. E. (2007). Confidence distribution (cd): distribution estimator of a parameter. *Lecture Notes-Monograph Series Vol. 54, Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, 54:132–150.

Smith, T., Spiegelhalter, D., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine*, 14(24):2685–2699.

Stettler, C., Wandel, S., Allemann, S., Kastrati, A., Morice, M. C., Schömig, A., Pfisterer, M. E., Stone, G. W., Leon, M. B., de Lezo, J. S., et al. (2007). Outcomes associated with drug-eluting and bare-metal stents: A collaborative network meta-analysis. *The Lancet*, 370(9591):937–948.

Sutton, A. J. and Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27(5):625–650.

Thorlund, K., Wetterslev, J., Awad, T., Thabane, L., and Gluud, C. (2011). Comparison of statistical inferences from the dersimonian–laird and alternative random-effects model meta-analyses–an empirical assessment of 920 cochrane primary outcome meta-analyses. *Research Synthesis Methods*, 2(4):238–253.

Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P.-Y., and Wei, L. J. (2009). Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2×2 tables with all available data but without artificial continuity correction. *Biostatistics*, 10(2):275–281.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter — a review (with discussion). *International Statistical Review*, 81(1):3–39.

Xie, M., Singh, K., and Strawderman, W. E. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association*, 106(493):320–333.