Statistics and Its Interface Volume 13 (2020) 425–436

On evidence cycles in network meta-analysis*

Lifeng Lin†, Haitao Chu‡, and James S. Hodges

As an extension of pairwise meta-analysis of two treatments, network meta-analysis has recently attracted many researchers in evidence-based medicine because it simultaneously synthesizes both direct and indirect evidence from multiple treatments and thus facilitates better decision making. The Bayesian hierarchical model is a popular method to implement network meta-analysis, and it is generally considered more powerful than conventional pairwise meta-analysis, leading to more precise effect estimates with narrower credible intervals. However, the improvement of effect estimates produced by Bayesian network meta-analysis has never been studied theoretically. This article shows that such improvement depends highly on evidence cycles in the treatment network. When all treatment comparisons are assumed to have different heterogeneity variances, a network meta-analysis produces posterior distributions identical to separate pairwise meta-analyses for treatment comparisons that are not contained in any evidence cycles. However, this equivalence does not hold under the commonly-used assumption of a common heterogeneity variance for all comparisons. Simulations and a case study are used to illustrate the equivalence of the Bayesian network and pairwise meta-analyses in certain networks.

Keywords and phrases: Bayesian hierarchical model, Evidence cycle, Indirect evidence, Network meta-analysis, Relative effect, Treatment network.

1. INTRODUCTION

Network meta-analysis of randomized controlled trials in evidence-based medicine, also known as mixed treatment comparison, has become an increasingly popular statistical method to simultaneously compare multiple treatments [40]. Based on an Internet Web search, the prestigious medical journals *BMJ*, *JAMA*, and *Lancet* have published more than 100 research articles with the term 'network meta-analysis' in their titles since 2010. A variety of methods are available for performing network meta-analysis [29, 38, 53, 52, 21, 12, 26, 20, 28, 25].

By comparing all treatments at a time, network meta-analysis can provide a coherent ranking of treatments and thus guide decision making [18, 41, 39]. More importantly, by synthesizing both direct and indirect evidence, network meta-analysis is generally considered more powerful than conventional pairwise meta-analysis, which compares each pair of treatments separately and thus can only use direct evidence [32]. For example, to compare treatments A and B, trials including both treatments provide direct evidence. Besides these trials, A and B can also be compared via a common comparator, say treatment C, and the trials of A vs. C and B vs. C provide indirect evidence. Such a trio of treatments forms an evidence cycle (or loop) [30]. For example, Shape 4 in Figure 1 contains one evidence cycle consisting of treatments 1, 2, and 5. Hence, a network meta-analysis is expected to produce more accurate effect estimates with narrower confidence/credible intervals (CI), compared with pairwise meta-analyses. For example, in a network meta-analysis of the efficacy of 12 new-generation antidepressants [5], the pairwise meta-analysis estimated the odds ratio of escitalopram vs. fluoxetine as 1.23 with 95% CI (0.87, 1.74), and the network meta-analysis estimated it as 1.32 with 95% CI (1.12, 1.55), indicating a statistically significant difference in efficacy between the two antidepressants.

Because of these attractive features, many researchers try to collect as many treatments as possible to enrich network meta-analyses, taking for granted the benefit from synthesizing direct and indirect evidence. Also, so far most methodological papers have been devoted to implementation and reporting of network meta-analyses, while researchers seldom carefully examine what we can gain from conducting network meta-analyses compared with much simpler pairwise meta-analyses [27]. The foregoing illustrative example of direct and indirect evidence of treatments A, B, and C is often used to introduce the idea of network meta-analysis. Based on this, it seems intuitive to conclude that evidence cycles are necessary for a network meta-analysis to outperform pairwise meta-analyses; otherwise, the two types of analyses must give identical results for treatment comparisons that are not in any evidence cycles.

*This research was supported in part by the U.S. National Institutes of Health NLM R21 012197 (HC, LL) and NLM R21 012744 (HC, JH). The content is solely the responsibility of the authors and does not necessarily represent official views of the National Institutes of Health. LL was also supported by the Doctoral Dissertation Fellowship from the University of Minnesota Graduate School. We thank the associate editor and two anonymous reviewers for helpful and constructive suggestions.

†Corresponding author. ORCID: 0000-0002-3562-9816.
‡ORCID: 0000-0003-0932-598X.

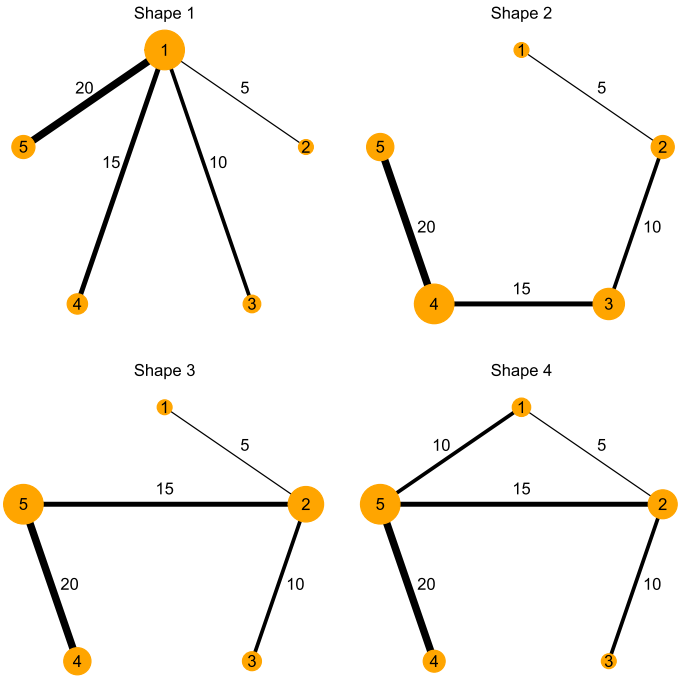

Figure 1. Artificial treatment networks with four shapes. Vertices represent treatments; edges represent direct comparisons. Edge width is proportional to the number of studies that report the corresponding direct comparison; vertex size is proportional to the number of studies that include the corresponding treatment.

However, results in published network meta-analyses are not always consistent with this intuition. For example, in a network meta-analysis of blood pressure-lowering agents in adults with diabetes and kidney disease for all-cause mortality [36], endothelin inhibitor was directly compared with placebo in two studies, and this comparison was not in any evidence cycles. A pairwise meta-analysis produced an odds ratio estimate 1.55 with 95% CI (0.82, 2.89) for this comparison, and a network meta-analysis led to 1.53 with 95% CI (0.79, 2.97); both results were similar. However, for another comparison between renin inhibitor and placebo that was also directly given in two studies and was not in any evidence cycles, pairwise and network meta-analyses estimated the odds ratio as 0.93 with 95% CI (0.39, 2.24) and 1.05 with 95% CI (0.81, 1.36), respectively. The point estimates were noticeably different and were in opposite directions, and their 95% CIs also differed dramatically.

In addition, some tutorials about network meta-analysis do not remind researchers to examine the differences between the results produced by pairwise and network meta-analyses, especially in the absence of evidence cycles. The National Institute for Health and Care Excellence (NICE) Decision Support Unit in the UK has provided excellent technical support documents on evidence synthesis using network meta-analyses. This series of tutorials has also been published in *Medical Decision Making*, and has set a benchmark for methods used in other work that NICE has undertaken [10]. However, in one tutorial (page 622 in Dias et al. [8]), a star-shaped network was presented as an example to illustrate network meta-analysis models. A star-shaped network refers to a network meta-analysis consisting of studies that share a common treatment (such as a placebo or a well-established standard treatment); see, e.g., Shape 1 in Figure 1. Moreover, the PRISMA-NMA statement [21] also provides guidance for reporting systematic reviews comparing multiple treatments using direct and indirect evidence, and has been widely used to prepare and report network meta-analyses. Like the NICE tutorials, the PRISMA-NMA statement uses some networks without evidence cycles to illustrate treatment networks (e.g., Figure 2 in Hutton et al. [21]). Although the statement correctly acknowledged that 'closed loops [cycles] are not required to be present for every comparison under study', it did not say what happens if some comparisons are not in any cycles. Using networks without evidence cycles in these tutorial papers might overstate the power of network meta-analyses. Possibly as a result, many network meta-analyses in the current literature do not contain cycles or include many comparisons that are not in cycles; see, e.g., Chatterjee et al. [3], Biondi-Zoccai et al. [1], Palmer et al. [36], Khera et al. [24], Chen et al. [4], and Tricco et al. [48]. Among the 186 network meta-analyses investigated by Nikolakopoulou et al. [35], 35 networks are star-shaped.

This article carefully explores how evidence cycles play a critical role in improving effect estimates produced by network meta-analysis compared with separate pairwise meta-analyses and when the two kinds of analysis give identical results. We focus on the network meta-analysis method using the Bayesian hierarchical model proposed by Lu and Ades [29], because it is currently the most widely used method; in the survey by Nikolakopoulou et al. [35], 111 out of 186 network meta-analyses used it. By showing the importance of evidence cycles, we remind researchers to be aware of situations in which network meta-analysis provides an advantage compared with separate pairwise meta-analyses. Our results provide some insight for journal editors and reviewers to evaluate future network meta-analyses.

This article is organized as follows. We first review the development of the Bayesian network meta-analysis model in Section 2. Section 3 shows theoretically that the joint posterior distributions of effect estimates produced by this network meta-analysis and by separate pairwise meta-analyses are identical for treatment comparisons without evidence cycles when assuming different heterogeneity variances for treatment comparisons. The proofs are in the Supplementary Materials. Simulations and a case study to illustrate the equivalence relationship are presented in Section 4, and Section 5 concludes with some remarks.

2. METHODS FOR GENERAL NETWORK META-ANALYSIS: A REVIEW

2.1 Bayesian hierarchical model for pairwise meta-analysis

Before reviewing the model for network meta-analysis, we begin with the Bayesian hierarchical model for the conventional pairwise meta-analysis proposed by Smith et al. [45], because it lays the foundation for the popular network meta-analysis model of Lu and Ades [29]. Suppose that a pairwise meta-analysis collects N studies and each study compares the same two treatments, such as an active treatment and a placebo. Let y_{i1} and y_{i2} be the observed aggregated outcome measures in study i's treatment groups 1 and 2, respectively. The overall relative effect comparing the two treatments is usually of interest. The random-effects pairwise meta-analysis model [45] can be generalized as follows to estimate the overall relative effect:

$$
\begin{aligned}
&y_{ik} \sim f(y \mid \Delta_{ik}, \xi_{ik}), \quad i = 1, \ldots, N, k = 1, 2; \\
&g(\Delta_{i1}) = \mu_i, \quad g(\Delta_{i2}) = \mu_i + \delta_i; \\
&\delta_i \sim N(d, \sigma^2).
\end{aligned} \tag{1}
$$

Here, μ_i is commonly called the baseline effect of study i and the study-specific relative effects δ_i are assumed to be exchangeable across studies with mean d, which is interpreted as the overall relative effect. The variance parameter σ^2 reflects heterogeneity between studies. The link function is $g(\cdot)$, and $f(\cdot \mid \cdot, \cdot)$ is the outcome measure's density function depending on an unknown location parameter Δ_{ik} and a nuisance parameter ξ_{ik}, which is assumed to be known. For example, if the outcome is continuous, y_{ik} is usually assumed to be normally distributed with unknown mean Δ_{ik} and known standard error ξ_{ik}, and $g(\cdot)$ is the identity link. If the outcome is binary, such as the condition of having a certain event, then y_{ik} is the number of events, which follows a binomial density with unknown event rate Δ_{ik} and known sample size ξ_{ik}. When the logit link function $\text{logit}(t) = \log[t/(1-t)]$ is used for binary outcomes, the fixed effect d represents the overall log odds ratio of treatment 2 compared to treatment 1.

2.2 Bayesian hierarchical model for network meta-analysis

Lu and Ades [29, 31] extended the pairwise meta-analysis model to multiple-treatment comparisons. Instead of comparing just two treatments, in a network meta-analysis N studies are included comparing a total of $K > 2$ treatments. Specifically, each study compares a subset of the K treatments; denote the treatment subset of study i as $\mathcal{T}_i$. A study is called a two-arm study if it compares two treatments, while a multi-arm study investigates more than two treatments. Again, assume that the observed aggregated outcome measure y_{ik} in study i's treatment group k follows the distribution $f(\cdot \mid \Delta_{ik}, \xi_{ik})$. To implement the network meta-analysis model, a baseline treatment b_i needs to be specified for each study i. Different studies can have different baseline treatments because the treatment subsets $\mathcal{T}_i$ need not intersect. We denote b_i simply as b when it does not lead to confusion. The random-effects model for network meta-analysis can be specified as follows:

$$
\begin{aligned}
&y_{ik} \sim f(y \mid \Delta_{ik}, \xi_{ik}), \quad i = 1, \ldots, N, k \in \mathcal{T}_i; \\
&g(\Delta_{ik}) = \mu_i + X_{ik}\delta_{ibk}; \\
&\delta_{ibk} \sim N(d_{bk}, \sigma^2_{bk}), \quad \text{Corr}(\delta_{ibh}, \delta_{ibk}) = \gamma_{bhk}, \quad h, k \in \mathcal{T}_i.
\end{aligned} \tag{2}
$$

Here, X_{ik} is a dummy variable; $X_{ik} = 0$ if $k = b$ and $X_{ik} = 1$ if $k \in \mathcal{T}_i \backslash \{b\}$. For a multi-arm study, the correlation between the treatment contrasts δ_{ibh} and δ_{ibk} is assumed to be γ_{bhk}. Again, μ_i represents the baseline effect of study i, the study-specific relative effects δ_{ibk} are assumed to be exchangeable, and we focus on estimating the overall relative effects of all treatment contrasts d_{hk} $(1 \le h \ne k \le K)$.

A critical assumption in network meta-analysis is the consistency equation for an evidence cycle, which consists of a trio of treatments, under which

$$
d_{hk} = d_{\ell k} - d_{\ell h}, \quad \text{for all } 1 \le h \ne k \ne \ell \le K. \tag{3}
$$

If a treatment network contains evidence cycles, this equation permits synthesis of direct and indirect evidence for the treatment comparisons in the cycles, so that the network meta-analysis borrows more information than a conventional pairwise meta-analysis, which uses only direct evidence. The consistency assumption may not hold in some cases, and alternative approaches have been proposed to deal with evidence inconsistency [30, 42, 9, 17, 51, 11]. For example, one can add inconsistency factors w to Equation (3), that is, $d_{hk} = d_{\ell k} - d_{\ell h} + w_{hk\ell}$. This method is closely related to the number of independent cycles in the network, which is quantified by the inconsistency degrees of freedom df_{IC} [30]. If all studies are two-armed, then $df_{\text{IC}} = T - K + 1$, where T is the number of all treatment comparisons, i.e., the edges in the network. However, when multi-arm studies are present, the definition of inconsistency degrees of freedom is fairly complex and needs to be considered case by case.

Besides random-effects models, fixed-effects models are also often used in meta-analysis. These models assume that the collected studies are homogeneous, that is, that the relative effects for each treatment comparison share a common mean across studies, and their variation is entirely due to sampling error within studies. To be specific, the fixed-effects model for pairwise meta-analysis is

$$
\begin{aligned}
&y_{ik} \sim f(y \mid \Delta_{ik}, \xi_{ik}), \quad i = 1, \ldots, N, k = 1, 2; \\
&g(\Delta_{i1}) = \mu_i, \quad g(\Delta_{i2}) = \mu_i + d,
\end{aligned} \tag{4}
$$

and the fixed-effects model for network meta-analysis is

$$(5)\qquad \begin{aligned} &y_{ik} \sim f(y \mid \Delta_{ik}, \xi_{ik}), \quad i = 1, \ldots, N, k \in \mathcal{T}_i; \\ &g(\Delta_{ik}) = \mu_i + X_{ik} d_{bk}. \end{aligned}$$

Implementation is easier for the fixed-effects model than the random-effects model because the latter involves complex specification of heterogeneity variances, which will be detailed in Section 3.4. However, the homogeneity assumption may be unrealistic in many cases [16, 33], and the CIs produced by the fixed-effects model may have low coverage probabilities if heterogeneity is present in some treatment comparisons [34].

3. TREATMENT COMPARISONS WITHOUT EVIDENCE CYCLES

3.1 Direct and indirect evidence in networks without cycles

Throughout this article, the treatment network is assumed to be connected; if the network consists of disjoint sub-networks, then a separate analysis can be applied to each sub-network. We first consider treatment networks without cycles; in such networks, all collected studies must be two-armed because multi-arm studies create evidence cycles. Consequently, we no longer need to account for the correlations between treatment contrasts within studies in the random-effects network meta-analysis model in Equation (2).

To investigate the performance of the network meta-analysis model for a network without cycles, we explore the posterior distributions of all treatment contrasts. The $(K-1)K/2$ treatment contrasts are denoted as a vector $\boldsymbol{e} = (d_{hk}; 1 \le h < k \le K)^T$. In graph theory, a connected network without cycles is a spanning tree and contains exactly $K-1$ edges; denote the set of these edges as a $(K-1)$-dimensional vector $\boldsymbol{e}_{\mathrm{b}} = (e_1, \ldots, e_{K-1})^T$, where $e_j = d_{hk}$ for some $h < k$ and each e_j provides direct evidence. Thus, the set of all treatment contrasts $\boldsymbol{e}$ can be split into two subsets: $\boldsymbol{e}_{\mathrm{b}}$, each contrast that is directly compared in the network, and a $(K-2)(K-1)/2$-dimensional vector $\boldsymbol{e}_{\mathrm{f}} = (d_{hk}; d_{hk} \notin \boldsymbol{e}_{\mathrm{b}})^T$ that can only be imputed from indirect evidence. Using the definition of Lu and Ades [30], the treatment contrasts in $\boldsymbol{e}_{\mathrm{b}}$ are basic parameters, which involve all K treatments but do not form cycles; those in $\boldsymbol{e}_{\mathrm{f}}$ are referred to as functional parameters because they can be represented as functions of the basic parameters. Evidence consistency as defined in Equation (3) cannot be checked for networks without cycles because evidence inconsistency occurs within cycles. Under the consistency assumption, $\boldsymbol{e}_{\mathrm{f}}$ is entirely determined by $\boldsymbol{e}_{\mathrm{b}}$; that is, we may write $\boldsymbol{e}_{\mathrm{f}} = \mathbf{A}\boldsymbol{e}_{\mathrm{b}}$, where $\mathbf{A}$ is a known $(K-2)(K-1)/2 \times (K-1)$ transformation matrix. We have the following proposition regarding the transformation matrix $\mathbf{A}$.

Proposition 1. *The transformation matrix* $\mathbf{A}$ *is unique for each set of basic parameters, and each entry of* $\mathbf{A}$ *is 0 or* ± 1.

Proposition 1 holds for any type of connected network, including those containing cycles, under the consistency assumption. In networks without cycles, there is only one set of basic parameters $\boldsymbol{e}_{\mathrm{b}}$, so the transformation matrix $\mathbf{A}$ is uniquely defined.

3.2 Equivalence of network meta-analysis and separate pairwise meta-analyses

In a network without cycles, suppose study i, which must be two-armed, compares treatments k_i vs. h_i $(h_i < k_i)$; that is, the corresponding treatment contrast is $d_{h_i k_i}$. For $j = 1, \ldots, K-1$, let $\mathcal{S}_j = \{i : d_{h_i k_i} = e_j\}$ be the set of studies that give the direct treatment comparison e_j. Consequently, the N studies $\mathcal{S} = \{1, \ldots, N\}$ in the network can be partitioned into $K-1$ subsets according to their treatment contrasts: $\mathcal{S} = \bigcup_{j=1}^{K-1} \mathcal{S}_j$. Moreover, let $\mathcal{D}_j = \{(y_{ik}, \xi_{ik}); i \in \mathcal{S}_j, k \in \mathcal{T}_i\}$ be the data (aggregated outcome measures and nuisance parameters) provided by the studies in $\mathcal{S}_j$ and let $\mathcal{D} = \bigcup_{j=1}^{K-1} \mathcal{D}_j$ be the full data in the whole network. The pairwise meta-analysis uses the data $\mathcal{D}_j$ for each j separately to estimate the corresponding treatment contrast e_j, and we denote the resulting posterior distribution as $p(e_j \mid \mathcal{D}_j)$. The network meta-analysis uses the full data $\mathcal{D}$ to simultaneously compare all treatments, and we denote the joint posterior distribution of the direct treatment contrasts as $p(\boldsymbol{e}_{\mathrm{b}} \mid \mathcal{D})$. We have the following proposition.

Proposition 2. *For a treatment network without evidence cycles, given the same set of priors, the fixed-effects network meta-analysis model (5) gives posterior distributions of direct treatment contrasts identical to those from separate fixed-effects pairwise meta-analysis model (4), that is,*

$$(6)\qquad p(\boldsymbol{e}_b \mid \mathcal{D}) = \prod_{j=1}^{K-1} p(e_j \mid \mathcal{D}_j).$$

This equation also holds for the random-effects pairwise and network meta-analyses in Equations (1) and (2), if the network meta-analysis model uses different heterogeneity variances for different treatment contrasts.

Equation (6) implies that the posterior distribution of e_j produced by the network meta-analysis model is only informed by the data in studies $\mathcal{S}_j$; thus, the posterior distributions of the e_j's are mutually independent.

Note that the decomposition of the joint posterior in Equation (6) does not hold in treatment networks with evidence cycles that are formed by different two-arm studies, because the cycles introduce posterior correlations between the e_j's. Even if a network contains a cycle that is entirely produced by a single multi-arm study, this decomposition may not hold. In that cycle, consider two treatment comparisons that share a common treatment arm. Because the

posteriors of both treatment comparisons use information from the same multi-arm study, they may be correlated.

Unlike the e_j's in $\boldsymbol{e}_\mathrm{b}$ that are directly compared in the network, the estimates of $\boldsymbol{e}_\mathrm{f}$ are entirely informed by indirect evidence. The network meta-analysis seems to be an efficient approach to simultaneously estimating all treatment contrasts, including the indirect ones. However, under the consistency assumption, the following proposition shows that separate pairwise meta-analysis models also produce posterior distributions of indirect treatment contrasts identical to those given by the network meta-analysis model.

Proposition 3. *Under the model settings in Proposition 2 and the assumption of evidence consistency, the joint posterior distributions of the indirect treatment contrasts $\boldsymbol{e}_f$ produced by the network meta-analysis model and by separate pairwise meta-analysis models are identical for a network without evidence cycles.*

Propositions 2 and 3 imply that for a network without cycles, the network meta-analysis does not change the posterior distributions (thus, point estimates and CIs) of any treatment contrasts produced by separate pairwise meta-analyses.

3.3 Acyclic treatment comparisons in general networks

In a general treatment network that may contain evidence cycles, it commonly occurs that some treatment comparisons are not in any cycles [43]; we refer to such treatment comparisons as acyclic comparisons. Proposition 2 can be extended to the posterior distributions of acyclic comparisons in networks with general shapes. Specifically, suppose that a network with K treatments contains J acyclic comparisons, denoted as $\boldsymbol{e}_\mathrm{a} = (e_1, \ldots, e_J)^T$.

Proposition 4. *For a network with K treatments, the number of acyclic comparisons J does not exceed $K-1$.*

Studies that report the acyclic comparison e_j ($j = 1, \ldots, J$) must be two-armed; otherwise, multi-arm studies create evidence cycles containing e_j, contradicting the definition of an acyclic comparison. As in Section 3.2, let $\mathcal{S}_j$ be the set of studies that report the acyclic comparison e_j, and $\mathcal{S}^\star = \mathcal{S} \backslash \bigcup_{j=1}^{J} \mathcal{S}_j$ be the studies that do not report acyclic comparisons. The studies in $\mathcal{S}^\star$ form a sub-network with evidence cycles if the set $\mathcal{S}^\star$ is not empty. Suppose that $\boldsymbol{e}_\mathrm{b}^\star$ is a set of basic parameters in the sub-network consisting of $\mathcal{S}^\star$; then $\boldsymbol{e}_\mathrm{b} = (\boldsymbol{e}_\mathrm{a}^T, \boldsymbol{e}_\mathrm{b}^{\star T})^T$ is a set of basic parameters for the full network $\mathcal{S}$. Also, denote the data provided by $\mathcal{S}_j$ as $\mathcal{D}_j$ and the data provided by $\mathcal{S}^\star$ as $\mathcal{D}^\star$. Then we have the following proposition.

Proposition 5. *For acyclic treatment comparisons in a general network, the network meta-analysis does not improve their posterior distributions compared with separate pairwise meta-analyses under the model settings in Proposition 2. Specifically, using the same set of priors in the two models, the joint posterior distribution of the basic parameters produced by the network meta-analysis model is*

$$p(\boldsymbol{e}_b \mid \mathcal{D}) = p(\boldsymbol{e}_b^\star \mid \mathcal{D}^\star) \prod_{j=1}^{J} p(e_j \mid \mathcal{D}_j). \tag{7}$$

Here, $p(\boldsymbol{e}_\mathrm{b} \mid \mathcal{D})$ is produced by the network meta-analysis on the full network $\mathcal{S}$, while $p(\boldsymbol{e}_\mathrm{b}^\star \mid \mathcal{D}^\star)$ is the posterior based on the sub-network consisting of $\mathcal{S}^\star$. In a network without cycles, the study set $\mathcal{S}^\star$ does not exist so that $p(\boldsymbol{e}_\mathrm{b}^\star \mid \mathcal{D}^\star)$ drops out of Equation (7), which is thus reduced to Equation (6). Proposition 5 can therefore be viewed as a generalization of Proposition 2. Note that the sub-network consisting of $\mathcal{S}^\star$ contains evidence cycles, so the evidence may be inconsistent; however, Proposition 5 still applies for this situation.

3.4 The random-effects network meta-analysis model with equal heterogeneity variances

We have shown theoretically the equivalence of pairwise and network meta-analyses for treatment comparisons that are not in evidence cycles. However, one may wonder why these analyses produced dramatically different estimates for the acyclic comparison of renin inhibitor vs. placebo in the network meta-analysis of blood pressure-lowering agents, discussed in the introduction section. This happened because many researchers reduce model complexity by using a simplified specification of the random-effects model with a common heterogeneity variance, instead of directly using the network meta-analysis model specified in Equation (2).

Modeling the heterogeneity variances and covariances is an important issue in the random-effects network meta-analysis model (2). The difficulty arises from the fundamental relationship of the relative effects, $\delta_{ihk} = \delta_{i\ell k} - \delta_{i\ell h}$, so the heterogeneity standard deviations are constrained by the triangular inequality

$$|\sigma_{\ell h} - \sigma_{\ell k}| \leq \sigma_{hk} \leq |\sigma_{\ell h} + \sigma_{\ell k}|. \tag{8}$$

In the history of network meta-analysis, treatment comparisons were originally assumed to share a common heterogeneity variance (i.e., $\sigma_{hk}^2 = \sigma^2$ for all h and k), so that the triangular inequality (8) trivially holds [19, 32, 29]. Also, the correlations between the treatment comparisons, i.e., γ_{bhk} in Equation (2), were assumed to be 0.5 for conceptual and technical simplicity. In fact, these assumptions were derived under conditions that are fairly unrealistic from an arm-based viewpoint: in all studies, treatment-specific underlying effects $\theta_{i1}, \ldots, \theta_{iK}$ are mutually independent with a common between-study variance τ^2; consequently,

$$\mathrm{Var}(\delta_{ihk}) = \mathrm{Var}(\theta_{ik} - \theta_{ih}) = 2\tau^2$$

and

$$\mathrm{Corr}(\delta_{i\ell k}, \delta_{i\ell h}) = \frac{\mathrm{Cov}(\theta_{ik} - \theta_{i\ell}, \theta_{ih} - \theta_{i\ell})}{\sqrt{2\tau^2 \cdot 2\tau^2}} = \frac{\mathrm{Var}(\theta_{i\ell})}{2\tau^2} = 0.5$$

for all $1 \leq h \neq k \neq \ell \leq K$. These assumptions have been widely applied in many applications (e.g., Cipriani et al. [5], Trelle et al. [47], and Khera et al. [24]), even though they impose a strong constraint on the treatment comparisons, which may be unrealistic [2]. Lu and Ades [31] relaxed these assumptions to allow different heterogeneity variances and correlations as in Equation (2), but this method has been seldom used so far because of its complexity.

Under the assumption of equal heterogeneity variances, the decomposition of the joint posterior distribution produced by the network meta-analysis in Equations (6) and (7) is no longer valid. We have the following proposition.

Proposition 6. *For a treatment network without evidence cycles, the random-effects network meta-analysis model (2) with equal heterogeneity variances $\sigma_{bk}^2 = \sigma^2$ is equivalent to simultaneously performing the random-effects pairwise meta-analyses (1) for studies $\mathcal{S}_j$ conditional on the common heterogeneity variance σ^2.*

The pairwise meta-analyses in Proposition 6 may not be deemed separate, because each pairwise meta-analysis uses the common heterogeneity variance σ^2, which is informed by all studies $\mathcal{S}$ instead of the study set $\mathcal{S}_j$ for a specific treatment contrast. Although the random-effects network meta-analysis can therefore produce different results from separate random-effects meta-analyses that have no constraints on the heterogeneity variances σ_{bk}^2, Proposition 6 implies that these differences are caused entirely by the specification of heterogeneity variances if the treatment network does not contain evidence cycles.

4. NUMERICAL STUDIES

4.1 Simulations

We conducted simulations to illustrate the important role of evidence cycles in network meta-analyses. The outcome was assumed to be continuous and normally distributed, and each treatment's outcome measure y_{ik} and its within-study standard error ξ_{ik} were observed. The situation of a binary outcome will be explored in a real data analysis in Section 4.2. We simulated data containing five treatments with four network shapes, shown in Figure 1. Shapes 1–3 do not contain cycles: Shape 1 is a star-shaped network with its center at treatment 1; Shape 2 is a chain-shaped network with treatment contrasts from 2 vs. 1 to 5 vs. 4; and Shape 3 is more general than the star and chain shapes. In each of Shapes 1–3, four treatment contrasts are observed and form a set of basic parameters $\boldsymbol{e}_{\mathrm{b}}$; these treatment contrasts are reported in 5, 10, 15, or 20 studies, as described in Figure 1, so each network contains a total of 50 studies. Compared with Shape 3, Shape 4 contains 10 further studies of treatment contrast 5 vs. 1, so treatments 1, 2, and 5 form an evidence cycle.

To simulate the outcome measures, we first generated samples for all five treatments in each study, and then omitted certain treatment arms to create networks with the shapes in Figure 1; the omitted data were assumed to be missing completely at random. Specifically, the five treatments' within-study standard errors were drawn from $\xi_{ik} \sim U(0.1, 1)$ for treatment k in study i. The observed treatment-specific outcome measure was generated from $y_{ik} \sim N(\theta_{ik}, \xi_{ik}^2)$, where θ_{ik} represents the underlying true measure of treatment effect. The study-specific true measures were drawn from $(\theta_{i1}, \ldots, \theta_{i5})^T \sim N((\theta_1, \ldots, \theta_5)^T, \boldsymbol{\Psi})$, where θ_k represents the overall mean of treatment k ($k = 1, \ldots, 5$), and $\boldsymbol{\Psi}$ represents the between-study variance-covariance matrix. Note that although the data were simulated based on treatment arms, they also met the assumptions in the meta-analysis models in Equations (1), (2), (4), and (5), which are contrast-based: the true relative effect of treatments k vs. h was $d_{hk} = \theta_k - \theta_h$ and the study-specific baseline effects followed normal distributions. We set $\theta_k = k$ and $\boldsymbol{\Psi} = \mathbf{DRD}$, where $\mathbf{R} = (\rho_{hk})$ was the correlation matrix with $\rho_{kk} = 1$ and $\rho_{hk} = 0.4$ ($1 \leq h \neq k \leq 5$), and the between-study standard deviations $\mathbf{D} = \mathrm{diag}(\tau_1, \ldots, \tau_5)$ were sampled in three cases: (i) all studies were homogeneous with $\tau_k = 0$; (ii) all treatments had a common heterogeneity standard deviation $\tau_k = \tau$ with $\tau \sim U(1, 1.5)$; and (iii) the five treatments had different heterogeneity standard deviations with $\tau_k \sim U(0.4k - 0.4, 0.4k)$ for $k = 1, \ldots, 5$. Finally, certain treatments in certain studies were randomly omitted to produce networks with Shapes 1–4. For example, in the network with Shape 1, treatments 3–5 were omitted in five studies, so these five studies compared treatments 2 vs. 1.

For each network shape, 1000 replicates of the network data were generated; for each replicate, the Markov chain Monte Carlo algorithm was applied to implement the network and pairwise meta-analyses using one chain, which contained a run of 50,000 iterations after a 20,000-update burn-in period. For both the network and pairwise meta-analyses, three model settings were considered: a fixed-effects model, a random-effects model with different heterogeneity variances, and a random-effects model with a common heterogeneity variance. Vague priors were used for the study-specific baseline effects and the basic parameters; $U(0, 10)$ priors were used for the heterogeneity standard deviations in the random-effects models. When the random-effects network meta-analysis with different heterogeneity variances was applied to the simulated data with Shape 4, due to the triangular inequality constraint (8), the prior of the heterogeneity standard deviation σ_{15} for the contrast 5 vs. 1 was set to $U(|\sigma_{12} - \sigma_{25}|, \sigma_{12} + \sigma_{25})$ as suggested by Lu and Ades [30]. The models' performance was evaluated according to biases and mean squared errors of the estimated relative effects and coverage probabilities of the 95% CIs.

Table 1 presents the results for some treatment contrasts for Case (iii) of the between-study standard deviations; the simulation results for Cases (i) and (ii) are in Tables S1 and S2 in the Supplementary Materials. Of note, the pairwise

Table 1. Biases (outside brackets), mean squared errors (inside parentheses), and 95% credible interval coverage probabilities (%, inside square brackets) of some estimated relative effects produced by network and pairwise meta-analyses in simulations. The data were simulated using different heterogeneity standard deviations for different treatment contrasts

Network shape	Treatment contrast	Network meta-analysis			Pairwise meta-analysis		
		FE	RE1	RE2	FE	RE1	RE2
Shape 1	d_{12}	−0.03	−0.03	−0.02	−0.03	−0.03	−0.02
		(0.21)	(0.20)	(0.20)	(0.21)	(0.20)	(0.20)
		[81]	[99]	[100]	[81]	[99]	[100]
	d_{15}	0.03	0.03	0.03	0.03	0.03	0.03
		(0.35)	(0.18)	(0.18)	(0.35)	(0.18)	(0.18)
		[41]	[96]	[92]	[40]	[95]	[92]
	d_{23}	0.02[a]	0.02	0.02	0.02[a]	0.02	NA
		(0.43)	(0.35)	(0.35)	(0.43)	(0.35)	
		[73]	[99]	[100]	[73]	[99]	
	d_{45}	0.04[a]	0.03	0.03	0.04[a]	0.03	NA
		(0.59[c])	(0.31)	(0.31)	(0.59[c])	(0.31)	
		[44]	[97]	[95]	[44]	[97]	
Shape 2	d_{12}	−0.02	−0.03	−0.02	−0.03	−0.03	−0.02
		(0.21)	(0.20)	(0.20)	(0.21)	(0.20)	(0.21)
		[82]	[99]	[100]	[81]	[99]	[100]
	d_{13}	−0.01[a]	−0.02	−0.01	−0.03[a]	−0.02	NA
		(0.42)	(0.35)	(0.36)	(0.42)	(0.35)	
		[76]	[99]	[100]	[75]	[99]	
	d_{15}	−0.01[b]	−0.03[a]	−0.02[a]	−0.03[b]	−0.02[a]	NA
		(1.09[d])	(0.73[d])	(0.74[d])	(1.09[d])	(0.73[d])	
		[63]	[99]	[98]	[62]	[99]	
	d_{45}	0.02	0.02	0.02	0.02	0.02	0.02
		(0.37)	(0.18)	(0.18)	(0.37)	(0.18)	(0.18)
		[37]	[96]	[92]	[36]	[96]	[92]
Shape 3	d_{12}	−0.02	−0.03	−0.03	−0.03	−0.03	−0.02
		(0.21)	(0.20)	(0.21)	(0.21)	(0.20)	(0.21)
		[82]	[99]	[100]	[81]	[99]	[100]
	d_{13}	−0.02[a]	−0.02	−0.02	−0.03[a]	−0.02	NA
		(0.42)	(0.35)	(0.37)	(0.42)	(0.35)	
		[75]	[99]	[100]	[75]	[99]	
	d_{15}	−0.01[a]	−0.02[a]	−0.02[a]	−0.02[a]	−0.02[a]	NA
		(0.63[c])	(0.44)	(0.44)	(0.63[c])	(0.43)	
		[62]	[99]	[99]	[61]	[99]	
	d_{45}	0.02	0.02	0.02	0.02	0.02	0.02
		(0.37)	(0.18)	(0.18)	(0.37)	(0.18)	(0.18)
		[37]	[96]	[92]	[36]	[96]	[92]
Shape 4	d_{12}	−0.03	−0.01	−0.01	−0.03	−0.03	−0.03
		(0.37)	(0.16)	(0.21)	(0.21)	(0.20)	(0.21)
		[51]	[97]	[98]	[81]	[99]	[100]
	d_{13}	−0.03[a]	−0.01	0.00	NA	NA	NA
		(0.56[c])	(0.31)	(0.37)			
		[58]	[97]	[98]			
	d_{15}	−0.01[a]	0.00	0.00	0.01[a]	0.01	0.00
		(0.41[c])	(0.21)	(0.23)	(0.63[d])	(0.38)	(0.38)
		[45]	[96]	[95]	[45]	[96]	[92]
	d_{45}	0.02	0.02	0.02	0.02	0.02	0.02
		(0.37)	(0.18)	(0.18)	(0.37)	(0.18)	(0.18)
		[36]	[96]	[93]	[36]	[96]	[93]

FE, fixed-effects model; RE1, random-effects model with different heterogeneity variances for different treatment contrasts; RE2, random-effects model with a common heterogeneity variance; NA, not applicable; d_{hk}, treatment k compared with h. Monte Carlo standard error of bias: a, 0.02–0.03; b, 0.03–0.04; otherwise, less than 0.02. Monte Carlo standard error of mean squared error: c, 0.02–0.03; d, 0.03–0.05; otherwise, less than 0.02. Monte Carlo standard errors of all coverage probabilities are less than 2 percentage points.

random-effects meta-analyses with a common heterogeneity variance could not use direct comparisons to impute the indirect comparisons d_{23} and d_{45} in the simulated networks with Shape 1 and d_{13} and d_{15} in those with Shapes 2 and 3, because the common heterogeneity variance introduced potential correlations between the results of the direct comparisons. Also, for the simulated networks with Shape 4, due to the presence of the evidence cycle consisting of treatments 1, 2, and 5, all three pairwise meta-analyses could not estimate the indirect comparison d_{13}. Consequently, the results of these comparisons are not in Table 1 and Tables S1 and S2.

Because the treatments were missing completely at random in all cases, each model produced nearly unbiased point estimates for each treatment contrast. In Case (i), where the treatment effects were homogeneous across studies, using either pairwise or network meta-analysis for all four networks in Figure 1, both the fixed- and random-effects models produced estimated relative effects with similar mean squared errors. Also, the fixed-effects model led to CI coverage probabilities that were fairly close to the nominal level 95%, while the two random-effects models produced slightly inflated coverage probabilities, indicating that their 95% CIs were wider than the fixed-effects model. However, in Cases (ii) and (iii), due to the presence of heterogeneity, the fixed-effects model led to very poor CI coverage probabilities, while those produced by the random-effects models were generally satisfactory; the mean squared errors produced by the fixed-effects model were also larger than those of the random-effects models. Moreover, in Case (iii), the true heterogeneity variances τ_k^2 differed across treatments, while the second random-effects model incorrectly assumed the τ_k^2's were equal. Interestingly, the results produced by this random-effects model were fairly similar to those produced by the correct random-effects model assuming different heterogeneity variances, although the incorrect model had a slightly low CI coverage probability for the treatment contrast d_{15} in the network with Shape 1 and d_{45} in the networks with Shapes 2–4. Most importantly, for the simulated networks with Shapes 1–3 without any evidence cycles, the pairwise meta-analysis produced effect estimates with biases, mean squared errors, and CI coverage probabilities almost identical to those produced by the network meta-analysis, with some slight differences due to Monte Carlo error. Therefore, the network meta-analysis did not improve the effect estimates compared with the pairwise meta-analysis, consistent with Propositions 2, 3, and 6. For the simulated networks with Shape 4 that contained one evidence cycle of treatments 1, 2, and 5, the results of the treatment contrast 5 vs. 4 (which was not in any evidence cycle) produced by the network and pairwise meta-analyses were almost identical, while the results of d_{12} and d_{15} (which were in the evidence cycle) produced by the network and pairwise meta-analyses were noticeably different. These results are consistent with Proposition 5.

4.2 Real data analysis

In addition to the simulations, we applied the network and pairwise meta-analyses to the data collected by Trikalinos et al. [49], consisting of 63 studies of four treatments for non-acute coronary artery disease. All studies were two-armed. We indexed the treatments as (1) medical therapy; (2) percutaneous transluminal balloon coronary angioplasty; (3) bare-metal stents; and (4) drug-eluting stents. The outcome was the number of deaths due to the disease in each treatment group, which was assumed to follow a binomial distribution. The complete data are available in Table S3 in the Supplementary Materials. We used the logit link function for the network and pairwise meta-analyses, so the overall relative effects produced by these models were log odds ratios comparing pairs among the four treatments. Also, in the network meta-analysis, the treatment with the smallest index in each study was used as the baseline.

Figure 2 presents the treatment network; we refer to this as the full network. The full network had one evidence cycle, while the treatment comparison 4 vs. 3 was acyclic as it was not contained in any cycles. To illustrate the performance of the network meta-analysis model in a network without evidence cycles, we removed the four studies that directly compared treatments 3 vs. 1 from the complete data; the remaining studies led to a chain-shaped network without cycles, which we call the reduced network. Network and pairwise meta-analyses were applied to both the full and reduced networks. In the network meta-analysis model, $\boldsymbol{e}_{\mathrm{b}} = (d_{12}, d_{23}, d_{34})^T$ was chosen as the set of basic parameters; thus, $\boldsymbol{e}_{\mathrm{f}} = (d_{13}, d_{14}, d_{24})^T$ was the set of functional parameters. As in Section 4.1's simulations, the three model settings were considered and vague normal priors were assigned to the study-specific baseline effects and the basic parameters. In the random-effects models, $U(0, 10)$ priors were used for the heterogeneity standard deviations σ_{12}, σ_{23}, and σ_{34}. When the random-effects network meta-analysis with

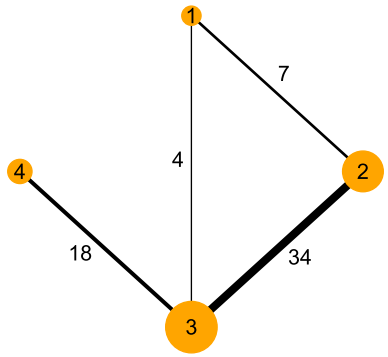

Figure 2. Network of four treatments on non-acute coronary artery disease. Treatment IDs: (1) medical therapy; (2) percutaneous transluminal balloon coronary angioplasty; (3) bare-metal stents; and (4) drug-eluting stents.

Table 2. Log odds ratios (95% credible intervals) comparing the four treatments for non-acute coronary artery disease

	Network meta-analysis			Pairwise meta-analysis		
LOR	FE	RE1	RE2	FE	RE1	RE2
Full network:						
d_{12}	−0.07	−0.16	−0.12	−0.21	−0.29	−0.29
	(−0.31, 0.17)	(−0.65, 0.32)	(−0.58, 0.28)	(−0.52, 0.09)	(−1.06, 0.30)	(−0.84, 0.20)
d_{13}	−0.11	−0.24	−0.22	−0.04	0.00	−0.01
	(−0.31, 0.08)	(−0.91[a], 0.25)	(−0.73, 0.20)	(−0.26, 0.18)	(−2.06[c], 2.64[c])	(−0.65, 0.73)
d_{14}	−0.03	−0.22	−0.19	NA	NA	NA
	(−0.49, 0.42)	(−1.10[a], 0.50[a])	(−0.98, 0.46)			
d_{23}	−0.05	−0.10	−0.10	−0.21	−0.22	−0.21
	(−0.29, 0.20)	(−0.58, 0.34)	(−0.47, 0.25)	(−0.53, 0.11)	(−0.81, 0.34)	(−0.62, 0.19)
d_{24}	0.03	−0.07	−0.07	NA	NA	NA
	(−0.45, 0.52)	(−0.83[a], 0.60)	(−0.75, 0.54)			
d_{34}	0.08	0.04	0.03	0.08	0.04	0.03
	(−0.33, 0.49)	(−0.56[a], 0.53)	(−0.52, 0.54)	(−0.33, 0.50)	(−0.55, 0.53)	(−0.53, 0.55)
Reduced chain-shaped network:						
d_{12}	−0.21	−0.29	−0.31	−0.21	−0.29	−0.30
	(−0.51, 0.09)	(−1.03[b], 0.31[a])	(−0.91, 0.24)	(−0.52, 0.09)	(−1.06, 0.30)	(−0.91, 0.24)
d_{13}	−0.42	−0.52	−0.52	−0.42	−0.53	−0.51
	(−0.86, 0.03)	(−1.41[b], 0.31[a])	(−1.26, 0.17)	(−0.86, 0.02)	(−1.47, 0.30)	(−1.26, 0.17)
d_{14}	−0.34	−0.49[a]	−0.51	−0.34	−0.49	−0.50
	(−0.95, 0.27)	(−1.55[b], 0.46[a])	(−1.48[a], 0.36)	(−0.95, 0.27)	(−1.60, 0.48)	(−1.48, 0.35)
d_{23}	−0.21	−0.22	−0.21	−0.21	−0.22	−0.21
	(−0.53, 0.11)	(−0.80, 0.34)	(−0.64, 0.21)	(−0.53, 0.11)	(−0.81, 0.34)	(−0.65, 0.21)
d_{24}	−0.13	−0.19	−0.20	−0.13	−0.18	−0.20
	(−0.66, 0.40)	(−1.00[a], 0.56)	(−0.94, 0.49)	(−0.66, 0.40)	(−1.02, 0.56)	(−0.94, 0.48)
d_{34}	0.08	0.04	0.01	0.08	0.04	0.01
	(−0.34, 0.50)	(−0.55[a], 0.53)	(−0.58, 0.55)	(−0.33, 0.50)	(−0.55, 0.53)	(−0.58, 0.55)

LOR, log odds ratio; FE, fixed-effects model; RE1, random-effects model with different heterogeneity variances for different treatment contrasts; RE2, random-effects model with a common heterogeneity variance; NA, not applicable; d_{hk}, treatment k compared with h. Monte Carlo standard error: a, 0.01–0.02; b, 0.02–0.03; c, 0.06–0.07; otherwise, less than 0.01.

different heterogeneity variances was applied to the full network, the prior of σ_{13} was set to $U(|\sigma_{12} - \sigma_{23}|, \sigma_{12} + \sigma_{23})$. Three chains were used to implement the network and pairwise meta-analyses via Markov chain Monte Carlo; each chain contained a run of 100,000 iterations after a 100,000-update burn-in period.

Table 2 presents the median overall log odds ratios of all treatment contrasts with their 95% CIs. When pairwise meta-analysis was applied to the full network, estimation of the indirect comparisons d_{14} and d_{24} was not possible without further assumptions due to unknown correlations between the separate estimated effects of d_{12}, d_{13}, d_{23}, and d_{34}; however, this problem was not present in the reduced network without cycles, as shown in Proposition 2. The potential scale reduction factors [14] of all traced parameters were much smaller than 1.05, indicating that the Markov chains had been stabilized; also, the convergence of the chains was checked using trace plots. In addition, we assessed the Monte Carlo standard errors of the point and interval estimates using the R package "mcmcse" [13] for the Markov chains. Most results had Monte Carlo standard errors much less than 0.01; those with standard errors greater than 0.01 are noted in Table 2.

For the reduced chain-shaped network, under each model setting, the network and pairwise meta-analyses produced nearly the same estimates of log odds ratios for all six treatment contrasts. Most differences between the two models were no more than 0.01 in absolute magnitude for point estimates and lower/upper bounds of 95% CIs, and they were due to Monte Carlo errors. These results were consistent with the propositions in Section 3. When the network meta-analysis was applied to the full network, Table 2 shows that the estimated overall log odds ratios of the basic parameters d_{12} and d_{23} differed from those using the reduced network; thus, d_{13}, d_{14}, and d_{24}, which were functions of the basic parameters d_{12} and d_{23}, also differed from their results using the reduced network. Recall that the reduced network only removed four studies that compared treatments 3 vs. 1. However, two of the four studies enrolled more than 1000 patients in each of their treatment groups, and they were the largest two among all 63 studies in the full network; see Table S3 in the Supplementary Materials. Thus, the removal of

these large studies caused the large differences noted above. Nevertheless, since the treatment contrast 4 vs. 3 was not contained in any evidence cycles, the estimated overall log odds ratio of d_{34} was nearly the same when the network meta-analysis was used for the full and reduced networks under both the fixed-effects setting and the random-effects setting with different heterogeneity variances. This is consistent with Proposition 5. When all treatment contrasts were assumed to have a common heterogeneity variance, the 95% CI of the log odds ratio for d_{34} using the reduced network slightly differed from that using the full network. This change arose because the estimate of d_{34} partly depended on the estimated heterogeneity variance, which was influenced by the removal from the full network of the four studies that compared treatments 3 vs. 1.

5. DISCUSSION

For treatment comparisons that are not in evidence cycles, the equivalence of pairwise and network meta-analyses is rarely observed in applications, even if the results from both types of meta-analyses are reported. As we have illustrated above, one reason is that most articles implemented the network meta-analysis model with a common heterogeneity variance for all treatment comparisons, while the separate pairwise meta-analyses used different heterogeneity variances for each comparison. Many researchers have adopted the assumption of a common heterogeneity variance in applications, but this assumption is seldom examined. This problem basically involves the tradeoff between goodness-of-fit and model complexity, which may be assessed using the deviance information criterion [46]. Assuming a common heterogeneity variance for all comparisons in the network meta-analysis may effectively reduce model complexity and avoid over-fitting. However, this assumption may lead to too-narrow 95% CIs with poor coverage, especially for some comparisons that are supported by only a few studies. Also, some network meta-analyses (e.g., Chen et al. [4]) contain many insufficiently-compared treatments; a comparison may be directly given by only one study and not be in any cycles. Because such a comparison is entirely informed by one source of direct evidence from a single study and there is nothing else to be synthesized, it may be inappropriate to perform a random-effects meta-analysis and make the strong assumption of a common heterogeneity variance.

In addition to the foregoing popular but possibly unrealistic assumption, using inconsistent analysis methods can also cause differences between pairwise and network meta-analyses in the absence of evidence cycles. Instead of the Bayesian hierarchical model in Equation (1) or (4), frequentist methods (e.g., the inverse-variance fixed-effect model, the DerSimonian–Laird random-effects model [7], or Hartung–Knapp–Sidik–Jonkman method [15, 44, 22, 23]) currently dominate pairwise meta-analyses, possibly because they can be easily implemented using various statistical software [50, 37]. For example, Cipriani et al. [6] conducted a Bayesian network meta-analysis using WinBUGS, while performing pairwise meta-analyses using frequentist methods in Stata. Therefore, when reporting results from both pairwise and network meta-analyses, researchers are encouraged to use consistent analysis methods, such as the Bayesian models in Equations (1) and (2) or frequentist pairwise and network meta-analysis methods [38]. By doing so, the benefit of network meta-analysis can be accurately reflected in the differences between the results from pairwise and network meta-analyses.

This article indicates that evidence cycles are critical for a network meta-analysis to improve the effect estimates and outperform separate pairwise meta-analyses. Such improvement depends strongly on the evidence consistency assumption (3) for each cycle, which effectively reduces the degrees of freedom of the total of $(K-1)K/2$ treatment comparisons d_{hk} $(1 \le h < k \le K)$. However, each cycle potentially suffers from evidence inconsistency [30, 11], which is caused by a discrepancy among the treatment comparisons within evidence cycles. By using inconsistency factors w for evidence cycles to deal with this problem, the degrees of freedom of the treatment contrasts increases, and the power of the network meta-analysis is accordingly reduced. In other words, researchers must accept a greater risk of evidence inconsistency to gain more power from a network meta-analysis.

In summary, the number of evidence cycles is a critical factor for journal editors and reviewers to evaluate network meta-analyses. Although the assumption of a common heterogeneity variance has been frequently used in practice, researchers should carefully examine this assumption because it may be unrealistic and lead to incorrect conclusions in some cases. Also, when performing a network meta-analysis using Bayesian methods, researchers should compare its results with those produced by Bayesian pairwise meta-analyses, instead of using frequentist pairwise meta-analysis methods.

SUPPLEMENTARY MATERIALS

Proofs of the propositions in Section 3, additional simulation results, and data for the case study in Section 4.2 are available in the Supplementary Materials (http://intlpress.com/site/pub/files/_supp/sii/2020/0013/0004/SII-2020-0013-0004-s001.pdf).

Received 2 November 2018

REFERENCES

[1] Biondi-Zoccai, G., Lotrionte, M., Thomsen, H. S., Romagnoli, E., D'Ascenzo, F., Giordano, A. and Frati, G. (2014). Nephropathy after administration of iso-osmolar and low-osmolar contrast media: evidence from a network meta-analysis. *International Journal of Cardiology* **172** 375–380.

[2] Brignardello-Petersen, R., Murad, M. H., Walter, S. D., McLeod, S., Carrasco-Labra, A., Rochwerg, B., Schünemann, H. J., Tomlinson, G. and Guyatt, G. H. (2019). GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. *Journal of Clinical Epidemiology* **105** 60–67.

[3] Chatterjee, S., Biondi-Zoccai, G., Abbate, A., D'Ascenzo, F., Castagno, D., Van Tassell, B., Mukherjee, D. and Lichstein, E. (2013). Benefits of β blockers in patients with heart failure and reduced ejection fraction: network meta-analysis. *BMJ* **346** f55.

[4] Chen, C., Zhang, X. and Ma, X. (2017). Durability of cervical disc arthroplasties and its influence factors: a systematic review and a network meta-analysis. *Medicine* **96** e5947.

[5] Cipriani, A., Furukawa, T. A., Salanti, G., Geddes, J. R., Higgins, J. P. T., Churchill, R., Watanabe, N., Nakagawa, A., Omori, I. M., McGuire, H., Tansella, M. and Barbui, C. (2009). Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *The Lancet* **373** 746–758.

[6] Cipriani, A., Zhou, X., Del Giovane, C., Hetrick, S. E., Qin, B., Whittington, C., Coghill, D., Zhang, Y., Hazell, P., Leucht, S., Cuijpers, P., Pu, J., Cohen, D., Ravindran, A. V., Liu, Y., Michael, K. D., Yang, L., Liu, L. and Xie, P. (2016). Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *The Lancet* **388** 881–890.

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

[8] Dias, S., Sutton, A. J., Welton, N. J. and Ades, A. E. (2013). Evidence synthesis for decision making 3: heterogeneity—subgroups, meta-regression, bias, and bias-adjustment. *Medical Decision Making* **33** 618–640.

[9] 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** 932–944. MR2752057

[10] Dias, S., Welton, N. J., Sutton, A. J. and Ades, A. E. (2013). Evidence synthesis for decision making 1: introduction. *Medical Decision Making* **33** 597–606.

[11] Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G. and Ades, A. E. (2013). Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Medical Decision Making* **33** 641–656.

[12] Efthimiou, O., Debray, T. P. A., van Valkenhoef, G., Trelle, S., Panayidou, K., Moons, K. G. M., Reitsma, J. B., Shang, A. and Salanti, G. (2016). GetReal in network meta-analysis: a review of the methodology. *Research Synthesis Methods* **7** 236–263.

[13] Flegal, J. M., Hughes, J., Vats, D. and Dai, N. (2017). mcmcse: Monte Carlo Standard Errors for MCMC R package version 1.3-2.

[14] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7** 457–472. MR1294072

[15] Hartung, J. and Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine* **20** 3875–3889.

[16] Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology* **37** 1158–1160.

[17] Higgins, J. P. T., Jackson, D., Barrett, J. K., Lu, G., Ades, A. E. and White, I. R. (2012). Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods* **3** 98–110.

[18] Higgins, J. P. T. and Welton, N. J. (2015). Network meta-analysis: a norm for comparative effectiveness? *The Lancet* **386** 628–630.

[19] Higgins, J. P. T. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **15** 2733–2749.

[20] Hong, H., Chu, H., Zhang, J. and Carlin, B. P. (2016). A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods* **7** 6–22.

[21] Hutton, B., Salanti, G., Caldwell, D. M., Chaimani, A., Schmid, C. H., Cameron, C., Ioannidis, J. P. A., Straus, S., Thorlund, K., Jansen, J. P., Mulrow, C., Catalá-López, F., Gøtzsche, P. C., Dickersin, K., Boutron, I., Altman, D. G. and Moher, D. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine* **162** 777–784.

[22] IntHout, J., Ioannidis, J. P. A. and Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology* **14** 25.

[23] Jackson, D., Law, M., Rücker, G. and Schwarzer, G. (2017). The Hartung-Knapp modification for random-effects meta-analysis: a useful refinement but are there any residual concerns? *Statistics in Medicine* **36** 3923–3934. MR3713639

[24] Khera, R., Murad, M. H., Chandar, A. K., Dulai, P. S., Wang, Z., Prokop, L. J., Loomba, R., Camilleri, M. and Singh, S. (2016). Association of pharmacological treatments for obesity with weight loss and adverse events: a systematic review and meta-analysis. *JAMA* **315** 2424–2434.

[25] Lin, L. (2018). Quantifying and presenting overall evidence in network meta-analysis. *Statistics in Medicine* **37** 4114–4125. MR3879417

[26] Lin, L., Chu, H. and Hodges, J. S. (2016). Sensitivity to excluding treatments in network meta-analysis. *Epidemiology* **27** 562–569.

[27] Lin, L., Xing, A., Kofler, M. J. and Murad, M. H. (2019). Borrowing of strength from indirect evidence in 40 network meta-analyses. *Journal of Clinical Epidemiology* **106** 41–49.

[28] Lin, L., Zhang, J., Hodges, J. S. and Chu, H. (2017). Performing arm-based network meta-analysis in R with the pcnetmeta package. *Journal of Statistical Software* **80** 1–25.

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

[30] Lu, G. and Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* **101** 447–459. MR2256166

[31] Lu, G. and Ades, A. E. (2009). Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* **10** 792–805.

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

[33] Ma, X., Lin, L., Qu, Z., Zhu, M. and Chu, H. (2018). Performance of between-study heterogeneity measures in the Cochrane Library. *Epidemiology* **29** 821–824.

[34] Mills, E. J., Thorlund, K. and Ioannidis, J. P. A. (2013). Demystifying trial networks and network meta-analysis. *BMJ* **346** f2914.

[35] Nikolakopoulou, A., Chaimani, A., Veroniki, A. A., Vasiliadis, H. S., Schmid, C. H. and Salanti, G. (2014). Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS ONE* **9** e86754.

[36] Palmer, S. C., Mavridis, D., Navarese, E., Craig, J. C., Tonelli, M., Salanti, G., Wiebe, N., Ruospo, M., Wheeler, D. C. and Strippoli, G. F. (2015). Comparative efficacy and safety of blood pressure-lowering agents in adults with diabetes and kidney disease: a network meta-analysis. *The Lancet* **385** 2047–2056.

[37] Palmer, T. M. and Sterne, J. A. C. (2016). *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*, 2nd ed. Stata Press, College Station, TX.

[38] Rücker, G. (2012). Network meta-analysis, electrical networks and graph theory. *Research Synthesis Methods* **3** 312–324.
[39] Rücker, G. and Schwarzer, G. (2015). Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology* **15** 58.
[40] Salanti, G. (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** 80–97.
[41] Salanti, G., Ades, A. E. and Ioannidis, J. P. A. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* **64** 163–171.
[42] Salanti, G., Higgins, J. P. T., Ades, A. E. and Ioannidis, J. P. A. (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* **17** 279–301. MR2521537
[43] Salanti, G., Kavvoura, F. K. and Ioannidis, J. P. A. (2008). Exploring the geometry of treatment networks. *Annals of Internal Medicine* **148** 544–553.
[44] Sidik, K. and Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in Medicine* **21** 3153–3159.
[45] Smith, T. C., Spiegelhalter, D. J. and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine* **14** 2685–2699.
[46] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** 583–639. MR1979380
[47] Trelle, S., Reichenbach, S., Wandel, S., Hildebrand, P., Tschannen, B., Villiger, P. M., Egger, M. and Jüni, P. (2011). Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. *BMJ* **342** c7086.
[48] Tricco, A. C., Zarin, W., Cardoso, R., Veroniki, A.-A., Khan, P. A., Nincic, V., Ghassemi, M., Warren, R., Sharpe, J. P., Page, A. V. and Straus, S. E. (2018). Efficacy, effectiveness, and safety of herpes zoster vaccines in adults aged 50 and older: systematic review and network meta-analysis. *BMJ* **363** k4029.
[49] Trikalinos, T. A., Alsheikh-Ali, A. A., Tatsioni, A., Nallamothu, B. K. and Kent, D. M. (2009). Percutaneous coronary interventions for non-acute coronary artery disease: a quantitative 20-year synopsis and a network meta-analysis. *The Lancet* **373** 911–918.
[50] Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36** 1–48.
[51] White, I. R., Barrett, J. K., Jackson, D. and Higgins, J. P. T. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* **3** 111–125.
[52] Yang, G., Liu, D., Liu, R. Y., Xie, M. and Hoaglin, D. C. (2014). Efficient network meta-analysis: a confidence distribution approach. *Statistical Methodology* **20** 105–125. MR3205725
[53] Zhang, J., Carlin, B. P., Neaton, J. D., Soon, G. G., Nie, L., Kane, R., Virnig, B. A. and Chu, H. (2014). Network meta-analysis of randomized clinical trials: reporting the proper summaries. *Clinical Trials* **11** 246–262.

Lifeng Lin
Department of Statistics,
Florida State University,
Tallahassee, FL, USA
E-mail address: linl@stat.fsu.edu

Haitao Chu
Division of Biostatistics,
University of Minnesota
Minneapolis, MN, USA
E-mail address: chux0051@umn.edu

James S. Hodges
Division of Biostatistics,
University of Minnesota
Minneapolis, MN, USA
E-mail address: hodge003@umn.edu