A Theoretical Investigation of How Evidence Flows in Bayesian Network Meta-Analysis of Disconnected Networks

Audrey Béliveau[*] and Paul Gustafson[†]

Abstract. Network meta-analysis has gained popularity in the last decade as a method for comparing the efficacy/safety of multiple medical interventions by synthesizing data across clinical studies. Bayesian methods for network meta-analysis have undergone further development than frequentist methods and are more convenient to use. Most of the current literature pertains to connected networks but disconnected networks commonly arise. There is not at the moment a trusted gold-standard approach to analyze disconnected networks. Intuitively, the standard method for analyzing connected networks, which is contrast-based, does not seem useful in disconnected networks, but this has not been explained rigorously. Our work is the first to provide the theoretical groundwork for understanding how evidence flows within Bayesian contrast-based models of disconnected networks. We achieve this by quantifying the ratio of posterior to prior variance of disconnected treatment contrasts. We show that when using an uninformative prior on the treatment contrasts, the standard approach is not useful to analyze disconnected networks (even when the number of studies, treatments or patients is large); however, it can be useful under moderately informative priors, which can be informed from additional observational data when available. A simulation study provides a demonstration of the theoretical results and explores non-asymptotic cases. An illustration on a real-world dataset is provided.

Keywords: Bayesian inference, Bayesian network meta-analysis, contrast-based approach, disconnected network, indirect evidence, mixed treatment comparison, network meta-analysis.

1 Introduction

Network meta-analysis (NMA) has become increasingly popular in the past decade for comparing the efficacy and/or safety of several medical interventions (Zarin et al., 2017). When developing a new treatment or intervention for a health condition, its efficacy/safety shall be compared with all the available evidence obtained through a systematic literature review of published studies. While meta-analysis only allows for the comparison of two treatments, network meta-analysis allows for the comparison of all relevant treatments in a single analysis in order to support policies and costs pertaining to newly developed treatments as well as guidelines for health practitioners

[*]Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada, audrey.beliveau@uwaterloo.ca
[†]Department of Statistics, University of British Columbia, 3182 Earth Sciences Building, 2207 Main Mall, Vancouver, British Columbia, V6T 1Z4, Canada, gustaf@stat.ubc.ca
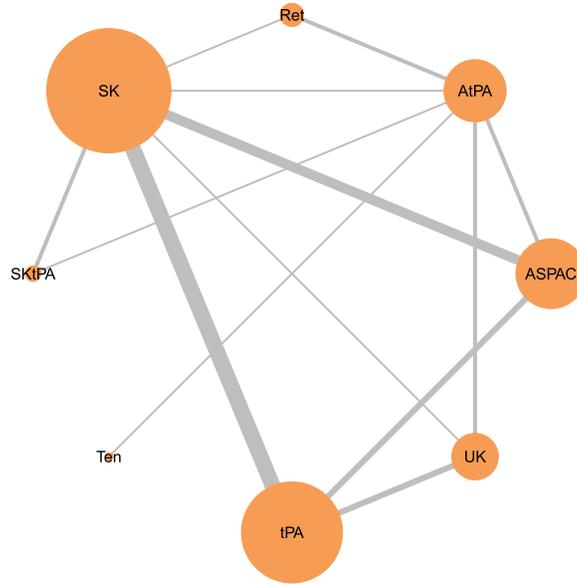
Figure 1: Schematic representation of a connected network of evidence. The size of the nodes and edges are proportional to the number of studies examining a specific treatment and the number of comparisons between two given treatments, respectively. Treatments are abbreviated as SK (streptokinase), Ret (reteplase), AtPA (accelerated alteplase), ASPAC (anistreplase), UK (urokinase), tPA (alteplase), Ten (tenecteplase) and SKtPA (streptokinase+alteplase); see Boland et al. (2003).

(Efthimiou et al., 2016). The use of a Bayesian framework is popular among practitioners for implementing network meta-analysis. Bayesian NMA methods have been developed to a further degree and are easier to implement than their frequentist counterparts (Hoaglin et al., 2011; Dias et al., 2013). They also have the advantage of lending themselves nicely to health-economic decision modeling (Dias et al., 2018).

For any outcome of interest, one can represent the NMA evidence base by drawing a network of nodes and edges, see Figure 1. The nodes represent the treatments while an edge between two nodes indicates that the treatments were compared in at least one study. For two given treatments, if there is a path in the network of evidence that connects the two treatments, we say that the treatments are connected; otherwise we say they are disconnected. Likewise if all pairs of treatments in the network are connected, we say that the network is connected (Figure 1); otherwise, we say that the network is disconnected (Figure 2). In other words, in a disconnected network there are at least two treatments, e.g. SK and AtPA, that have neither been compared directly within a study (SK vs AtPA) nor are related through comparator treatments over a series of studies (e.g. SK vs Ret and Ret vs AtPA).

Standard NMA methods were developed for connected networks (Lu and Ades, 2004) but in recent years there has been a noticeable interest in disconnected networks
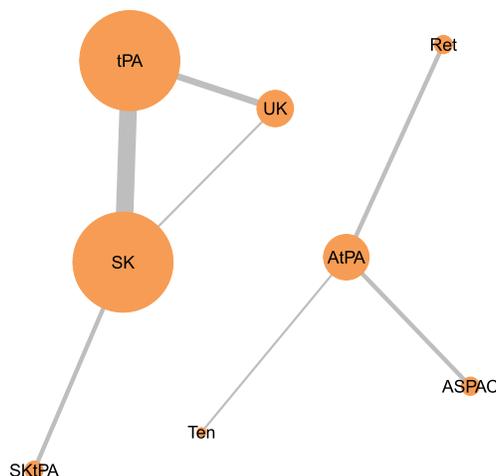
Figure 2: Schematic representation of a disconnected network of evidence. The size of the nodes and edges are proportional to the number of studies examining a specific treatment and the number of comparisons between two given treatments, respectively.

(Stevens et al., 2018; Thom et al., 2019; Rucker et al., 2019; Schmitz et al., 2018). In practice, many situations can lead to disconnected networks. An example is when, for a given set of studies, a primary outcome of interest (e.g. remission) produces a connected network of evidence but other secondary outcomes of interest (e.g. side effects) were not measured or reported in all studies. This results in disconnected networks for those secondary outcomes. Another example is when outcomes are measured on different scales, patient populations or at different follow-up times or when different dosages were used across studies. This likely results in a disconnected network unless a connected network is formed by pooling scales, follow-up times and/or drug dosages which may not be reasonable. Goring et al. (2016) provide other examples of contexts where disconnected networks tend to occur such as when there is no accepted standard of care, when the use of the standard of care is debated, when there are several accepted standards of care or when there has been a major change in standard of care.

The data for an NMA is usually obtained at the aggregate level rather than the individual patient level, because the latter is typically confidential. In connected networks, this aggregate data is commonly analyzed using a contrast-based model following Lu and Ades (2004); Dias et al. (2013). The overwhelmingly common choice is to use independent priors on the trial-specific baseline treatment effects – not treating the latter as exchangeable is the traditionally accepted practice in meta-analysis (Higgins et al., 2019; Dias and Ades, 2016). However one may use common parameters across studies to model the study-specific relative differences between treatments. Thus in essence we are deliberately choosing to allow borrowing strength across studies about relative difference between treatments and deliberately choosing to not borrow strength across studies about the absolute risk of the outcome. This has been seen to work very sensibly in connected networks, enabling the desired indirect comparisons of pairs of treatment

that have never been implemented in the same trial. However, practitioners have had to explore alternative methods for the analysis of disconnected networks as the standard approach would not yield useful inference for disconnected contrasts. An illustration of the failure of the standard model can be found in a small simulation study by Goring et al. (2016) where the posterior distribution on contrasts of disconnected treatments was not updated significantly from the prior distribution.

Alternative statistical methods have been proposed for the analysis of disconnected networks but they rely on strong assumptions. Stevens et al. (2018) provide a systematic review of methods which includes: use of external controls, treatment effect parameter, random baseline models, adjusted treatment response (including external evidence-based adjustment, iterative proportional fitting, propensity score matching methods, unanchored matching adjusted indirect comparisons (MAIC), simulated treatment comparisons (STC)), model-based meta-analysis, multivariate meta-analysis and class effect models. Among those methods, MAIC and STC have been used in health technology assessment (HTA) submissions but have been consistently criticized (Pooley et al., 2017). These methods require individual patient-level data for one of the studies and require that all the effects modifiers and prognostics factors are known and available (Signorovitch et al., 2010, 2012; Caro and Ishak, 2010; Ishak et al. , 2015; Phillippo et al., 2018, 2016). The later assumption is, as reported by the National Institute for Health and Care Excellence (NICE), widely accepted as "very hard to meet" and "unanchored comparisons based on disconnected networks and/or involving single-arm studies are therefore problematic" (Phillippo et al., 2016). As pointed out by Phillippo et al. (2018), "further research is needed to assess all available methods alongside MAIC and STC; in particular, to examine their properties and robustness to breakdown of assumptions, with varying levels of data availability, through thorough simulation studies." Overall, the analysis of disconnected networks is still an area of controversy and further theoretical and empirical investigations are needed to better understand the properties of the statistical methods proposed.

With this paper, we introduce for the first time a theoretical framework to understand how evidence flows in Bayesian models of disconnected networks with respect to estimating the relative effect of disconnected treatments. We achieve this by considering the standard contrast-based NMA model (Dias et al., 2013) and evaluating theoretically, in disconnected networks, the ratio of posterior to prior variance between all pairs of disconnected treatments. If treatments A and B are not connected, would we literally expect no evidence about the relative effect of treatment B compared to A, in the sense that its posterior variance is the same as its prior variance? Or might there be some reduction in posterior variance, and if so how much? Moreover, might such a reduction depend on the number of studies, number of treatments, sample sizes and/or choice of priors?

In Section 2 we describe the Bayesian contrast-based model used through this paper. In Section 3, we describe our approach and methodology for the theoretical work. Through theoretical developments in Section 4, we show that the ratio of posterior to prior variance for contrasts of disconnected treatments is bounded below by 0.5 when using a standard uninformative prior (provided some asymptotic conditions are met). The standard contrast-based NMA model is therefore not useful to analyze disconnected

networks (even when the number of studies, treatments or patients is large) when using uninformative priors. A lower bound is also obtained for an alternate form of uninformative prior and similar conclusions are reached. In Section 5, we conduct a simulation study which gives an empirical confirmation of the theoretical results. In the simulation study, we also explore the use of moderately informative priors on the treatment contrasts and find that, contrary to uninformative priors, they allow evidence to flow in the network to estimate disconnected contrasts. We also explore the case of small sample sizes (15 per arm) and find that our theoretical lower bounds are quite robust. In Section 6 we illustrate our results on a real-world dataset and find that our theoretical lower bounds on the posterior to prior variance ratio are quite robust to the small sample sizes and small number of events. Finally, we conclude with a discussion in Section 7.

2 Setup

Suppose we have s studies (or trials) labeled $1, \ldots, s$ and t treatments labeled $1, \ldots, t$. Assume without loss of generality that treatment 1 is identified as the baseline treatment in each study (which could be missing in some studies). The purpose of the baseline treatment will be clarified when formulating the model.

In any trial $i \in \{1, \ldots, s\}$, suppose we have $a_i \geq 2$ treatment arms (groups), labeled $1, \ldots, a_i$. Note that we exclude the possibility of single-arm studies as their incorporation in NMA is not straightforward and often not recommended due to the lack of a comparator and randomization. Suppose without loss of generality that arms are labeled after being sorted in ascending order by treatment label. Let A_i be the $a_i \times (t-1)$ design matrix for each trial i, defined such that $A_{i,kj} = 1$ if the k-th arm involves treatment $j + 1$, otherwise $A_{i,kj} = 0$. Thus, each row-sum is either zero or one, depending on whether the arm involves treatment 1 or not.

The binomial data is given by n_{ik} and R_{ik} and is available for arms $k \in \{1, \ldots, a_i\}$ of studies $i \in \{1, \ldots, s\}$. The former, n_{ik}, denotes the number of individuals in the k-th arm of study i. The latter, R_{ik}, is a binomial outcome, i.e. a count of the number of individuals that responded favorably (or unfavorably) to the treatment in the k-th arm of study i. Note that although we focus on a binomial outcome through the paper, theory could be developed analogously for other types of outcomes, such as rate, count or continuous outcomes. This would simply involve adapting the normal approximation and delta-method presented at the beginning of Section 4 to accommodate different distributions and link functions.

2.1 Bayesian contrast-based model

The generally accepted approach for the network meta-analysis of connected networks is the contrast-based (CB) approach (Lu and Ades, 2004; Dias et al., 2013). Nowadays, it is overwhelmingly implemented in a Bayesian way, which we follow in this paper.

When R_{ik} is a binomial outcome, the natural way to model such data is to start with a binomial distribution, that is:

$$R_{ik}|\phi_{ik} \overset{\text{independent}}{\sim} \text{Binomial}(n_{ik}, \ \phi_{ik}), \tag{2.1}$$

where the parameter $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{ia_i})^\top$ is further modeled using a logit link:

$$\text{logit}(\boldsymbol{\phi}_i) = \mathbf{1}_{a_i}\alpha_i + \text{A}_i\boldsymbol{\delta}_i, \tag{2.2}$$

where $\mathbf{1}_{a_i}$ is a vector of ones of length a_i, α_i is the absolute effect (log-odds) of treatment 1 in trial i and $\boldsymbol{\delta}_i = (\delta_{i2}, \ldots, \delta_{it})^\top$, where δ_{ij} is the effect of treatment j relative to treatment 1 in trial i (log-odds ratio). Note that α_i is defined even for studies that do not include treatment 1 (the baseline treatment). Also, δ_{ij} is defined even if treatment j is not included in study i and the multiplication by A_i in (2.2) selects the appropriate treatments. The use of the within-study relative treatment effects $\boldsymbol{\delta}_i$ is key to the contrast-based approach. It is assumed that the within-study relative treatment effect of any two treatments j and $j' \neq j$ can be expressed through the transitivity relation $\delta_{i,jj'} \equiv \delta_{ij'} - \delta_{ij}$ (Dias et al., 2013).

Strength is borrowed across studies when modeling the study-specific relative effects between treatments:

$$\boldsymbol{\delta}_i|\mathbf{d}, \Sigma \sim \text{N}(\mathbf{d}, \Sigma), \tag{2.3}$$

with mean relative treatment effects $\mathbf{d} = (d_{12}, \ldots, d_{1t})^\top$. This use of a common $\mathbf{d}$ across studies enables the comparisons of treatments that have never been implemented in the same trial by allowing evidence to flow in the network. For Σ we adopt the usual compound symmetry structure described in Dias et al. (2013), with variances τ^2 and covariances $0.5\tau^2$, i.e. $\Sigma = (\tau^2/2)(\text{J}_{t-1} + \text{I}_{t-1})$ with I_{t-1}, the $(t-1) \times (t-1)$ identity matrix, and J_{t-1}, the $(t-1) \times (t-1)$ matrix of ones. The compound symmetry structure ensures that study-specific contrasts have the same variance for any pair of treatments. This way, the conditional variance of $\delta_{i,jj'} = \delta_{ij'} - \delta_{ij}$, for any $j' \neq j$ and the conditional variance of δ_{ij} are the same, that is τ^2. Independent priors are used on τ^2 and $\mathbf{d}$.

The aim of any network meta-analysis is typically to compare treatments. The posterior distribution on the parameter $\mathbf{d}$ permits inference on contrasts with treatment 1 only, the baseline treatment. To compare other pairs of treatments (j, j') where $j \neq j' \neq 1$, we compute the posterior distribution of $d_{jj'} \equiv d_{1j'} - d_{1j}$, which follows from the transitivity of the δ_{ij}'s.

Strength is not borrowed across trials to inform α_i, the absolute log-odds of the outcome in study i for the baseline treatment. From a frequentist perspective, this is equivalent to saying the α_i's are fixed effects. As such, independent priors are used on the α_i's: $\alpha_i|\mu_\alpha, \sigma_\alpha^2 \overset{iid}{\sim} N(\mu_\alpha, \sigma_\alpha^2)$ for set values of μ_α and σ_α^2. For example, to obtain an uninformative prior, one can take $\mu_\alpha = 0$ and a large value for σ_α^2.

3   Methodology

Suppose we observe data as described in Section 2 that forms a disconnected network where treatments $1, \ldots, t^*$ are disconnected from treatments $t^*+1, \ldots, t$, with $\min(t^*, t-t^*) \geq 2$. We say that treatments $t^* + 1, \ldots, t$ form the auxiliary network, which is disconnected from the main network composed of treatments $1, \ldots, t^*$. Let $p = t - t^*$ be the number of treatments in the auxiliary network.

Suppose the observed data for this disconnected network are analyzed using the contrast-based model presented in Section 2.1. We want to assess theoretically whether in disconnected networks the posterior distribution of $d_{jj'}$ for disconnected treatments j and j' (i.e. $j \in \{1, \ldots, t^*\}$ and $j' \in \{t^* + 1, \ldots, t\}$) can be updated significantly from the prior distribution. To do so, we consider an asymptotic framework (large enough sample size in each study) and derive theoretical results that can apply to any number of studies, treatments or patients. In cases where we find that the posterior variance of $d_{jj'}$ does not differ significantly from the prior variance, for any disconnected treatments j and j', we would have demonstrated theoretically that the Bayesian contrast-based approach lacks utility in those cases.

We consider two possible priors on $\mathbf{d}$, which we denote respectively by π_1 and π_2. Both have the multivariate normal form $\mathbf{d} \sim N(\mathbf{0}, \Sigma_d)$. For π_1, we use a variance-covariance matrix proportional to the identity, that is: $\Sigma_d = \sigma_d^2 I_{t-1}$, where σ_d^2 takes a numerical value. This choice of prior is recommended by the National Institute for Health and Care Excellence (Dias et al., 2013). With this prior, the prior variance of contrasts that involve the baseline treatment is σ_d^2 while it is $2\sigma_d^2$ for other contrasts, which might not be desirable. Thus for π_2, we use $\Sigma_d = \sigma_d^2(I_{t-1} + J_{t-1})/2$, which has the desirable property of equal prior variance (σ_d^2) on each contrast. Note that π_2 imposes a correlation of 0.5 between elements of $\mathbf{d}$ whereas they are independent using π_1.

We do not impose a structure on the prior for τ^2 as this is not necessary for the theoretical developments. For the prior on the α_i's, we do not prescribe a value for μ_α but ensure that the prior is uninformative (which is the common practice) by taking the limit $\sigma_\alpha^2 \to \infty$.

A reader not interested in the theoretical derivations may skip to Table 1 for a summary of the theoretical results.

4 Theoretical developments on the posterior to prior variance ratio of disconnected contrasts

In order to gain theoretical insight, we start by approximating the responses with a normal distribution. Assuming the n_{ik}'s are large enough, a normal approximation on (2.1) is

$$\frac{R_{ik}}{n_{ik}} \mid \phi_{ik} \approx N[\phi_{ik}, \ \phi_{ik}(1 - \phi_{ik})/n_{ik}]. \tag{4.1}$$

Furthermore, setting $Y_{ik} = \text{logit}(R_{ik}/n_{ik})$ for all i's and k's and applying the delta-method, we get that

$$\mathbf{Y}_i \mid \boldsymbol{\phi}_i \sim N[\text{logit}(\boldsymbol{\phi}_i), \text{diag}(\mathbf{v}_i)], \tag{4.2}$$

where $v_{ik} = \{n_{ik}\phi_{ik}(1 - \phi_{ik})\}^{-1}$, and $\text{logit}(\boldsymbol{\phi}_i)$ is modeled as in Section 2 as $\text{logit}(\boldsymbol{\phi}_i) = \mathbf{1}_{a_i}\alpha_i + A_i\boldsymbol{\delta}_i$ with $\alpha_i \stackrel{iid}{\sim} N(\mu_\alpha, \sigma_\alpha^2)$ and $\boldsymbol{\delta}_i \mid \mathbf{d}, \tau^2 \stackrel{iid}{\sim} N\{\mathbf{d}, (\tau^2/2)(I_{t-1} + J_{t-1})\}$. We take the v_{ik}'s as known to make mathematical derivations possible, following standard meta-analytic derivations, see Borenstein et al. (2009) chapters 5 and 14. The assumptions necessary for the normal approximation (4.1) to hold ensure that the ν_{ik}'s can

be approximated precisely from the data and thus they can be treated as fixed. Fixed values for the variance of the normal distribution also arise naturally when the outcome of interest is continuous (as opposed to binomial). Means and standard error estimates are extracted from a systematic literature review and the standard error estimates are then treated as *fixed* quantities in the variance of a normal model.

Marginally, $f(\mathbf{Y}_i \mid \mathbf{d}, \tau^2)$ is normally distributed and the mean and variance parameters can be obtained using the laws of total expectation and variance as $\mathrm{E}\{\mathrm{E}(\mathbf{Y}_i \mid \boldsymbol{\delta}_i, \alpha_i) \mid \mathbf{d}, \tau^2\}$ and $\mathrm{E}\{\mathrm{var}(\mathbf{Y}_i \mid \boldsymbol{\delta}_i, \alpha_i) \mid \mathbf{d}, \tau^2\} + \mathrm{var}\{\mathrm{E}(\mathbf{Y}_i \mid \boldsymbol{\delta}_i, \alpha_i) \mid \mathbf{d}, \tau^2\}$. This gives $\mathbf{Y}_i \mid \mathbf{d}, \tau^2 \sim N\left(\mathbf{1}_{a_i}\mu_\alpha + \mathrm{A}_i\mathbf{d}, G_i\right)$, where $G_i = \mathrm{diag}(\mathbf{v}_i) + \sigma_\alpha^2 \mathrm{J}_{a_i} + (\tau^2/2)\mathrm{A}_i(\mathrm{I}_{t-1} + \mathrm{J}_{t-1})\mathrm{A}_i^T$.

Since we are interested in the posterior distribution of $\mathbf{d}$ we now use the Bayes theorem $\pi(\mathbf{d} \mid \mathrm{Y}, \tau^2) \propto f(\mathrm{Y} \mid \mathbf{d}, \tau^2)\pi(\mathbf{d})$. Assuming a multivariate normal prior on $\mathbf{d}$, of the form $\mathbf{d} \sim N(\mathbf{0}, \Sigma_d)$ we get

$$\mathrm{var}(\mathbf{d} \mid \mathrm{Y}, \tau^2) = \left(\Sigma_d^{-1} + \Delta_{\sigma_\alpha}\right)^{-1}, \tag{4.3}$$

where $\Delta_{\sigma_\alpha} = \sum_{i=1}^{s} \mathrm{A}_i^\top G_i^{-1} \mathrm{A}_i$. We develop Δ_{σ_α} using the definition of A_i and take the limit of interest $\sigma_\alpha^2 \to \infty$. The details of those calculations and a general expression for $\Delta = \lim_{\sigma_\alpha^2 \to \infty} \Delta_{\sigma_\alpha}$ are in Online Appendix A (Béliveau and Gustafson, 2020a). It is important to point out that Δ is symmetric and block-diagonal of the form

$$\Delta = \begin{pmatrix} \Delta_A & 0 \\ 0 & \Delta_B \end{pmatrix},$$

where Δ_A is a $(t^* - 1) \times (t^* - 1)$ matrix and Δ_B is a $(t - t^*) \times (t - t^*)$ singular matrix. Both matrices Δ_A and Δ_B are functions of τ^2, the $\mathbf{v}_i$'s and the design matrices A_i.

Now we exploit the block-diagonal form of Δ. Let $V^* \equiv \mathrm{var}(\mathbf{d} \mid \mathrm{Y}, \tau^2)$ and $P \equiv \Sigma_d^{-1}$, then

$$\begin{pmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{pmatrix} = \begin{pmatrix} P_{11} + \Delta_A & P_{12} \\ P_{21} & P_{22} + \Delta_B \end{pmatrix}^{-1}. \tag{4.4}$$

Hence, $V_{22}^* = \left\{P_{22} + \Delta_B - P_{21}(P_{11} + \Delta_A)^{-1}P_{12}\right\}^{-1}$. We are interested in the value of V_{22}^* because it corresponds to the posterior distribution of disconnected contrasts between treatments in the auxiliary network and the baseline treatment, conditional on τ^2 (we will eliminate the conditioning on τ^2 later).

Theorem 4.1. *When the prior variance on* $\mathbf{d}$ *is* π_1, *we have*

$$V_{22}^* = (P_{22} + \Delta_B)^{-1}, \tag{4.5}$$

where $P_{22} = I_{t-t^*}/\sigma_d^2$.

Proof. This is obtained simply by noting that when the prior variance on $\mathbf{d}$ is π_1, we have $P_{21} = P_{12} = 0$. □

Theorem 4.2. *When the prior variance on* $\mathbf{d}$ *is* π_2, *and assuming that* σ_d^2 *is large (relative to the elements of* Δ_A, *which do not depend on* σ_d^2) *we have*

$$V_{22}^* \approx (P_{22} + \Delta_B)^{-1}, \tag{4.6}$$

where $P_{22} = 2(I_{t-t^*} - J_{t-t^*}/t)/\sigma_d^2$.

Proof. When the prior variance on $\mathbf{d}$ is π_2, we have $P_{12} = P_{21}^\top = -2J_{(t^*-1)\times(t-t^*)}/(t\sigma_d^2)$. Because we do not have $P_{21} = P_{12} = 0$, things do not develop as nicely. However, assuming that σ_d^2 is large (relative to the elements of Δ_A, which do not depend on σ_d^2) we have that $P_{21}(P_{11} + \Delta_A)^{-1}P_{12} \approx P_{21}\Delta_A^{-1}P_{12}$ is negligible relative to P_{22}. Thus, $V_{22}^* \approx (P_{22} + \Delta_B)^{-1}$, where $P_{22} = 2(I_{t-t^*} - J_{t-t^*}/t)/\sigma_d^2$. $\qquad\square$

The diagonal elements of V_{22}^* are of particular interest because they are the conditional posterior variances of disconnected contrasts between treatments in the auxiliary network and the baseline treatment, i.e. the $\text{var}(d_{1j'} \mid Y, \tau^2)$'s. Theorems 4.3 and 4.4 develop the value of V_{22}^* further and consider how it behaves in the non informative prior limit of $\sigma_d^2 \to \infty$ (and the point mass prior limit of σ_d^2).

Theorem 4.3. *Under prior $\pi_1(\mathbf{d})$ we have $\lim_{\sigma_d^2 \to \infty} \text{var}(d_{1j'} \mid Y, \tau^2)/\sigma_d^2 = 1/p$ and $\lim_{\sigma_d^2 \to 0} \text{var}(d_{1j'} \mid Y, \tau^2)/\sigma_d^2 = 1$ for any treatment $j' \neq 1$ in the auxiliary network.*

The proof of Theorem 4.3 is given in Online Appendix B.

Theorem 4.4. *Under prior $\pi_2(\mathbf{d})$ we have $\lim_{\sigma_d^2 \to \infty} \text{var}(d_{1j'} \mid Y, \tau^2)/\sigma_d^2 = 0.5p^{-1}/(1 - pt^{-1})$ for any treatment $j' \neq 1$ in the auxiliary network.*

The proof of Theorem 4.4 is given in Online Appendix C.

The results obtained in Theorems 4.3 and 4.4 concern uniquely disconnected contrasts that involve treatment 1 but we are interested in all treatment contrasts.

Theorem 4.5. *For any treatments j from the main network and j' from the auxiliary network, $\text{var}(d_{jj'} \mid Y, \tau^2) \geq \text{var}(d_{1j'} \mid Y, \tau^2)$ under prior $\pi_1(\mathbf{d})$. The same result holds approximately under prior $\pi_2(\mathbf{d})$ when σ_d^2 is large.*

Proof. First, we note that $\text{var}(d_{jj'} \mid Y, \tau^2) = \text{var}(d_{1j} \mid Y, \tau^2) + \text{var}(d_{1j'} \mid Y, \tau^2) - 2\text{Cov}(d_{1j}, d_{1j'} \mid Y, \tau^2) \geq \text{var}(d_{1j'} \mid Y, \tau^2) - 2\text{Cov}(d_{1j}, d_{1j'} \mid Y, \tau^2)$. In the case of prior $\pi_1(\mathbf{d})$, we have $P_{12} = P_{21} = 0$, thus $V_{12}^* = V_{21}^* = 0$. As a result, $\text{Cov}(d_{1j}, d_{1j'} \mid Y, \tau^2) = 0$ and we can conclude $\text{var}(d_{jj'} \mid Y, \tau^2) \geq \text{var}(d_{1j'} \mid Y, \tau^2)$. In the case of prior $\pi_2(\mathbf{d})$, we have when σ_d^2 is large, $P_{12} = P_{21} \approx 0$ thus $\text{var}(d_{jj'} \mid Y, \tau^2) \geq \text{var}(d_{1j'} \mid Y, \tau^2)$ also holds approximately. $\qquad\square$

Application of Fatou's lemma reveals that Theorems 4.3 and 4.4 limits are lower bounds when replacing the conditional posterior variances $\text{var}(. \mid Y, \tau^2)$ for their unconditional counterparts $\text{var}(. \mid Y)$. Thus, for any treatments j and j' in the main and auxiliary network, respectively, we have under prior $\pi_1(\mathbf{d})$,

$$\lim_{\sigma_d^2 \to \infty} \text{var}(d_{jj'} \mid Y)/\sigma_d^2 \geq p^{-1}, \tag{4.7}$$

and

$$\lim_{\sigma_d^2 \to 0} \text{var}(d_{jj'} \mid Y)/\sigma_d^2 \geq 1; \tag{4.8}$$

and under prior $\pi_2(\mathbf{d})$,

$$\lim_{\sigma_d^2 \to \infty} \text{var}(d_{jj'} \mid Y)/\sigma_d^2 \geq 0.5p^{-1}/(1 - pt^{-1}). \tag{4.9}$$

Taking into account the fact that under prior π_1 when $j \neq 1$ the prior variance of $d_{jj'}$ is $2\sigma_d^2$ rather than σ_d^2 (see Section 3), we construct Table 1 to summarize our theoretical results on the ratio of posterior to prior variance for disconnected treatment contrasts. Row 1 is based on the results of 4.3 and 4.4. Row 3 is based on the results of Theorem 4.5. Finally, rows 2 and 4 are based on (4.7)–(4.9).

Row id	Variance ratio	$\sigma_d^2 \to \infty$		$\sigma_d^2 \to 0$	
		$\pi_1(\mathbf{d})$	$\pi_2(\mathbf{d})$	$\pi_1(\mathbf{d})$	$\pi_2(\mathbf{d})$
1	$\frac{\mathrm{var}(d_{1j'}\mid Y, \tau^2)}{\mathrm{var}(d_{1j'})}$	p^{-1}	$0.5p^{-1}/(1 - pt^{-1})$	1	N/A
2	$\frac{\mathrm{var}(d_{1j'}\mid Y)}{\mathrm{var}(d_{1j'})}$	$\geq p^{-1}$	$\geq 0.5p^{-1}/(1 - pt^{-1})$	≥ 1	N/A
3	$\frac{\mathrm{var}(d_{jj'}\mid Y, \tau^2)}{\mathrm{var}(d_{jj'})}$	$\geq 0.5p^{-1}$	$\geq 0.5p^{-1}/(1 - pt^{-1})$	≥ 0.5	N/A
4	$\frac{\mathrm{var}(d_{jj'}\mid Y)}{\mathrm{var}(d_{jj'})}$	$\geq 0.5p^{-1}$	$\geq 0.5p^{-1}/(1 - pt^{-1})$	≥ 0.5	N/A

Table 1: Value or lower bound of the posterior-to-prior variance ratio specified on the left under the conditions specified at the top. Treatment j represents any treatment from the main network that is not treatment 1, and j' is any treatment from the auxiliary network (j and j' are disconnected). Note: t is the number of treatments in the network and p is the number of treatments in the auxiliary network.

4.1   Discussion of the theoretical results

Firstly from Table 1 we see that when using the contrast-based approach to analyze data from a disconnected network, with either prior π_1 or π_2 and an overly large value of σ_d^2, the data do not inform the posterior distribution of $\mathbf{d}$ significantly for disconnected contrasts. Because the prior variance is very large, even if the posterior distribution was reduced from the prior variance by a factor of 2, 10 or 100 the posterior distribution would remain very wide and negligibly informed by the data. This holds true even when increasing the number of studies or the number of patients because the results only depend on p and t. However in practice, depending on the choice of a "large" σ_d^2, a factor of 10 or 100 could be moderately useful. This will be explored with a simulation study in Section 5.

Secondly from Table 1 we see that under an informative point mass prior ($\sigma_d^2 \to 0$) of the form π_1, the data could potentially inform the posterior variance significantly for disconnected contrasts but only for those not involving treatment 1. In those case, the ratio of posterior to prior variance could be as low as 0.5. This is an interesting result which may be an artifact of using a prior where contrasts not involving treatment 1 were given twice the variance of contrasts involving treatment 1. This will be explored in the simulation study of Section 5.

Thirdly, in the special case of Online Appendix B.1 (two treatments in the auxiliary subnetwork, and prior π_1) we were able to show that the posterior variance is a weighted average of two quantities and thus bounded above and below. In that case, we can say with certitude that $0.5 \leq \mathrm{var}(d_{1j'} \mid Y)/\sigma_d^2 \leq 1$, for any value of σ_d^2. For all other cases, our asymptotic results do not allow us to conclude on the size of $\mathrm{var}(d_{1j'} \mid Y)$ for a

moderately informative prior on **d** but this will be explored in the simulation study of Section 5.

5   Simulation study

We conducted a small simulation study to attend to the following five objectives:

O1: Give an empirical demonstration of (4.3) and (A.1) (Online Appendix).

O2: Give an empirical demonstration of the theoretical results in rows 1 and 3 of Table 1.

O3: Give an empirical demonstration of the theoretical results in rows 2 and 4 of Table 1.

O4: Explore what happens when the sample size is large but moderately informative priors are used (we do not have general theoretical results for this).

O5: Explore a scenario with small sample sizes (we do not have theoretical results for this).

We considered a simple design with 5 treatments, labeled 1 to 5, and 3 studies. In the first study patients were randomized to treatments 1 and 2; in the second study patients were randomized to treatments 2 and 3; and in the third study patients were randomized to treatments 4 and 5. Hence, treatments 1, 2 and 3 form the main network, with treatment 1 as the baseline treatment; and treatments 4 and 5 form the auxiliary network. Schematically, this network is presented in Figure 3. We considered two scenarios with this design: 1000 patients per arm and 15 patients per arm. The former will be used to assess objectives O1 to O4 (large-sample sizes were assumed to derive the theory). The second scenario addresses objective O5.

We generated one dataset per scenario using the design (we do not generate many datasets per scenario as our goal in this work is not to assess the average performance of a statistical method over a large number of datasets). To generate the datasets, we generated the d_{1j}'s independently using $N(0, 0.15^2)$ distributions and then generated the
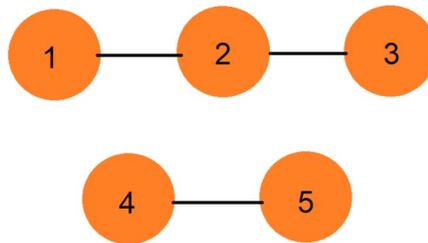

Figure 3: Schematic representation of the network used in the simulation study.

δ_{ij}'s using (2.3) with $\tau^2 = 0.01$. We generated the α_i's independently using $N(0.5, 0.5^2)$ distributions. We then generated binomial data, with 1000 (scenario 1) or 15 (scenario 2) patients per arms, using (2.1) and (2.2).

For the analysis of both datasets, we considered both priors $\pi_1(\mathbf{d})$ and $\pi_2(\mathbf{d})$ and considered a range of values for σ_d^2: 0.001 (very informative – nearly a point mass prior), 2 (moderately informative), 5 (moderately informative) and 100 (uninformative). Note that the magnitude of σ_d^2 shall be interpreted on the logit scale.

To address objectives O1 to O5, we analyzed the datasets using three methods, which we call M1, M2 and M3 for convenience. Note that in practice only method M3 can be used. Methods M1 and M2 can be implemented in the context of this simulation study because all parameters are known. For each method, we compute the ratio of posterior to prior variance on the treatment contrasts.

Method M1 consists of using formula (4.3) to calculate $\mathrm{var}(d_{jj'} \mid \mathrm{Y}, \tau^2)/\sigma_d^2$ for all connected and disconnected contrasts while substituting Δ_{σ_α} by Δ from (A.1) (Online Appendix) and setting τ^2 at its true value, 0.01.

Method M2 is an implementation of the contrast-based model described in Section 2 using the JAGS software (v. 4.3.0), except that we pretend that $\tau^2 = 0.01$ is known. For the α_i's we used independent $N(0, 100^2)$ priors. For method M2, we used a sample of 50,000 iterations thinned from 3 chains ran until convergence, which was assessed with traceplots. Initial values were generated randomly, and we used 10,000 adaptations and 100,000 burn-in iterations. The run times to convergence ranged from 0.4 mins with $\sigma_d^2 = 0.001$ to 24 mins with $\sigma_d^2 = 100$ for 3 chains in parallel on an Intel Core i7-8650U processor. Traceplots and R code are provided in Online Appendix and R Code Supplement (Béliveau and Gustafson, 2020b), respectively.

Method M3 is an implementation of the contrast-based model described in Section 2 using the JAGS software. Contrary to method M2, τ^2 is not fixed with method M3 and a Unif$(0, 20)$ prior is used on τ^2. For the α_i's we use independent $N(0, 100^2)$ priors as in method M2. We used the same MCMC settings as method M2 and run times were similar. Traceplots and R code are provided in Online Appendix and R Code Supplement, respectively.

To address objective O1, we compare the results between methods M1 and M2 in the $n = 1000$ scenario; if our theory is correct, methods M2 and M1 should give very similar values of $\mathrm{var}(d_{jj'} \mid \mathrm{Y}, \tau^2)$. Objective O2 is addressed by comparing values of $\mathrm{var}(d_{jj'} \mid \mathrm{Y}, \tau^2)$ from disconnected contrasts for method M2 in the $n = 1000$ scenario with rows 1 and 3 from Table 1. Objective O3 is addressed by comparing values of $\mathrm{var}(d_{jj'} \mid \mathrm{Y})$ from disconnected contrasts for method M3 in the $n = 1000$ scenario with rows 2 and 4 from Table 1. Objective O4 is addressed by looking at the results from method M3 when $\sigma_d^2 = 2$ or 5 in the $n = 1000$ scenario. Objective O5 is addressed by looking at the results from all methods in the $n = 15$ scenario.

The results for scenarios 1 ($n = 1000$) and 2 ($n = 15$) are presented in Figure 4 and 5 respectively.
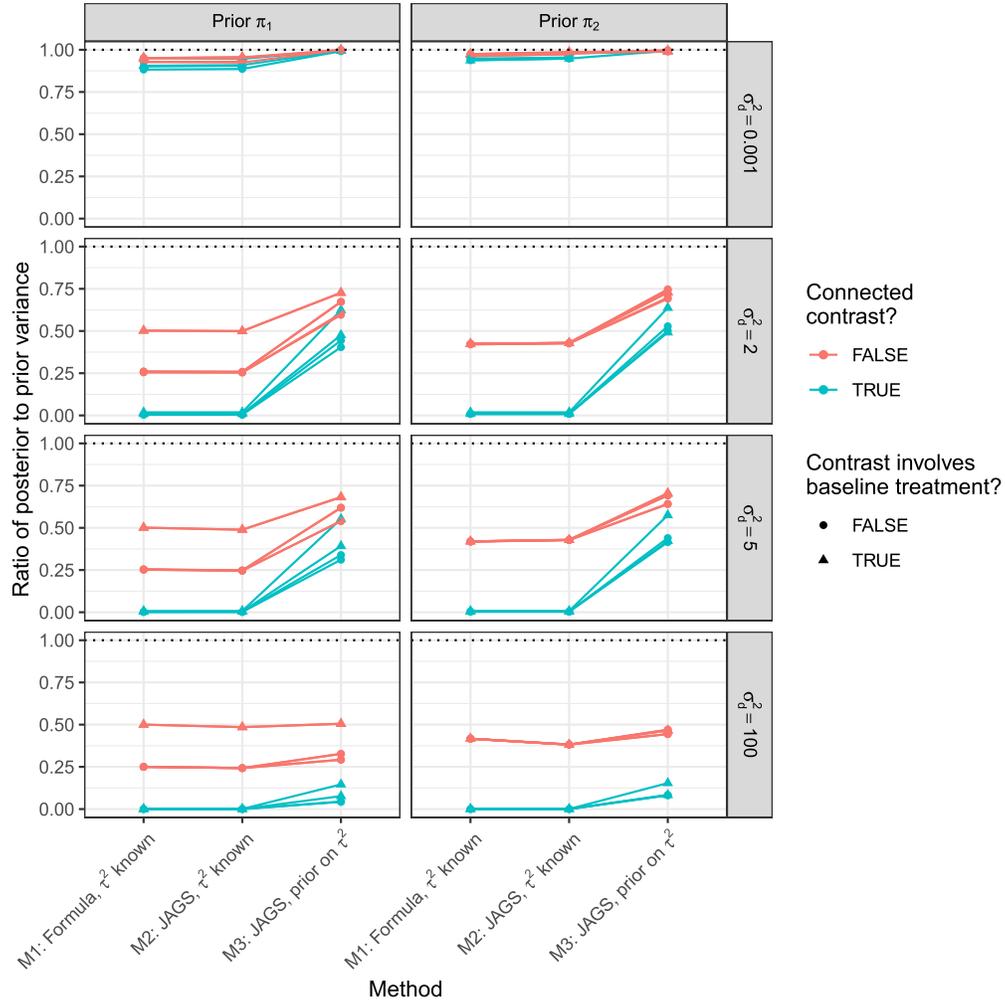
Figure 4: Results for the simulation study on scenario 1, using 1000 patients per arm. This graph shows the posterior-to-prior variance ratio for all (10) treatment contrasts obtained by methods M1, M2 and M3, with priors π_1 and π_2 and four different values of σ_d^2. Note that some points and lines appear stacked on top of each other. The dotted horizontal lines indicate the value 1 corresponding to no difference between the posterior and the prior variance.

5.1 Objective O1

In Figure 4 we see that when σ_d^2 is equal to 0.001, 2, 5 or 100 the results from methods M1 and M2 are in line with each other, which supports the correctness of (4.3) and (A.1) (Online Appendix). The largest discrepancy between M1 and M2 is observed

when $\sigma_d^2 = 100$ and this could be explained by the convergence not being as good for this scenario despite the longest run time.

5.2   Objective O2

For this objective, we look at the results from disconnected contrasts for method M2 in Figure 4.

When σ_d^2 is equal to 100 (uninformative prior on **d**), all disconnected contrasts (red) that involve the baseline treatment (triangle) have a ratio of posterior to prior variance close to 0.5 for prior π_1 and 0.42 for prior π_2. Moreover, all disconnected contrasts (red) that do not involve the baseline treatment (circle) have a ratio of posterior to prior variance close to 0.25 for prior π_1 and 0.42 for prior π_2. These numbers are in line with the theoretical results obtained when $\sigma_d^2 \to \infty$ summarized in rows 1 and 3 of Table 1. Note that here $p^{-1} = 0.5$ and $0.5p^{-1}/(1 - pt^{-1}) = 0.42$.

When $\sigma_d^2 = 0.001$ (very informative – close to point mass – prior on **d**) and prior π_1 is used, the results for method M2 are close to 1, although slightly lower. This shows support towards our theoretical results obtained when $\sigma_d^2 \to 0$ summarized in rows 1 and 3 of Table 1. The fact that the lower bound is not quite met could be explained by the fact that the limit $\sigma_d^2 \to 0$ is not attained.

5.3   Objective O3

For this objective, we look at the results from disconnected contrasts for method M3 in Figure 4.

When $\sigma_d^2 = 100$ (uninformative prior on **d**), all disconnected contrasts (red) that involve the baseline treatment (triangle) have a ratio of posterior to prior variance close to the theoretical lower bounds 0.5 for prior π_1 and 0.42 for prior π_2. Moreover, all disconnected contrasts (red) that do not involve the baseline treatment (circle) have a ratio of posterior to prior variance close to the theoretical lower bounds 0.25 for prior π_1 and 0.42 for prior π_2. These numbers are in line with the theoretical results obtained when $\sigma_d^2 \to \infty$ summarized in rows 1 and 3 of Table 1.

When $\sigma_d^2 = 0.001$ (very informative – close to point mass – prior on **d**) and prior π_1 is used, the ratio of posterior to prior variance with method M3 is very close to 1 for all contrasts. This result is consistent with the theoretical results obtained when $\sigma_d^2 \to 0$ summarized in rows 2 and 4 of Table 1.

5.4   Objective O4

For this objective, we look at the results from disconnected contrasts for method M3 in Figure 4.

When σ_d^2 is equal 2 or 5 (moderately informative priors on **d**) all disconnected contrasts (red) have a ratio of posterior to prior variance ≥ 0.5 for prior π_1 and ≥ 0.42

for prior π_2. These numbers are in line with the theoretical lower bounds obtained when $\sigma_d^2 \to \infty$ summarized in rows 2 and 4 of Table 1, even though $\sigma_d^2 = 2$ or 5 is not large.

Disconnected contrasts show ratios of posterior to prior variance between 0.5 to 0.75 which are meaningful reductions in prior variances for prior variances of moderate magnitudes. Some disconnected contrasts even show a ratio of posterior to prior variance smaller than another connected contrast from the same network.

Under prior π_1, we do not see a posterior to prior variance ratio twice as large in contrasts that involve the baseline treatment compared to contrasts not involving the baseline treatment (despite the prior variance on contrasts not involving the baseline treatment being twice as large as the prior variance on contrasts involving the baseline treatment). This can be explained by the fact that here the priors are informative: there might not be enough information in the data to counteract the prior information. In practice, it would be important that the choice of a moderately informative prior for **d** reflects accurately the information available a priori on each of the contrasts. This likely would not take the forms π_1 nor π_2.

5.5 Objective O5

The results from scenario 2 are presented in Figure 5. Despite the use of 15 patients per arm the asymptotics hold very well as methods M1 and M2 give the same results. The theoretical lower bounds from Table 1 for the ratio of posterior to prior variance of disconnected contrasts are satisfied across all M3 analyses even with a sample size of 15. When an uninformative prior is used ($\sigma_d^2 = 100$), the standard contrast-based approach does not allow disconnected contrasts to be estimated with sufficient precision. However, when moderately informative priors are used ($\sigma_d^2 = 2$ or 5), disconnected contrasts show ratios of posterior to prior variance between 0.55 to 0.8 which are meaningful reductions in prior variances when the for prior variances of moderate magnitudes. Some disconnected contrasts even have a ratio of posterior to prior variance smaller than another connected contrast from the same network.

6 Application to thrombolysis

We consider a dataset of 28 studies examining the effects of 8 thrombolytic drugs on mortality within 30 to 35 days (binary outcome) among individuals who have had an acute myocardial infarction (Boland et al., 2003). From the connected thrombolysis dataset (Figure 1), we generate a disconnected dataset arbitrarily, with the treatments SK, SKtPA, tPA and UK disconnected from the treatments ASPAC, AtPA, Ret and Ten (Figure 2). We achieve this by dropping all the two arm studies that compared any of the treatments that ought to be disconnected. We also drop one arm in each of the three arm studies to achieve the desired disconnections. The disconnected dataset contains 19 studies ranging from 59 to 20,163 patients. Event rates vary from 2–11% across arms with 29% of arms having no more than 5 events.

We analyze the disconnected dataset with the BUGSnet R package for contrast-based Bayesian NMA (Béliveau et al., 2019) using a binomial family distribution and
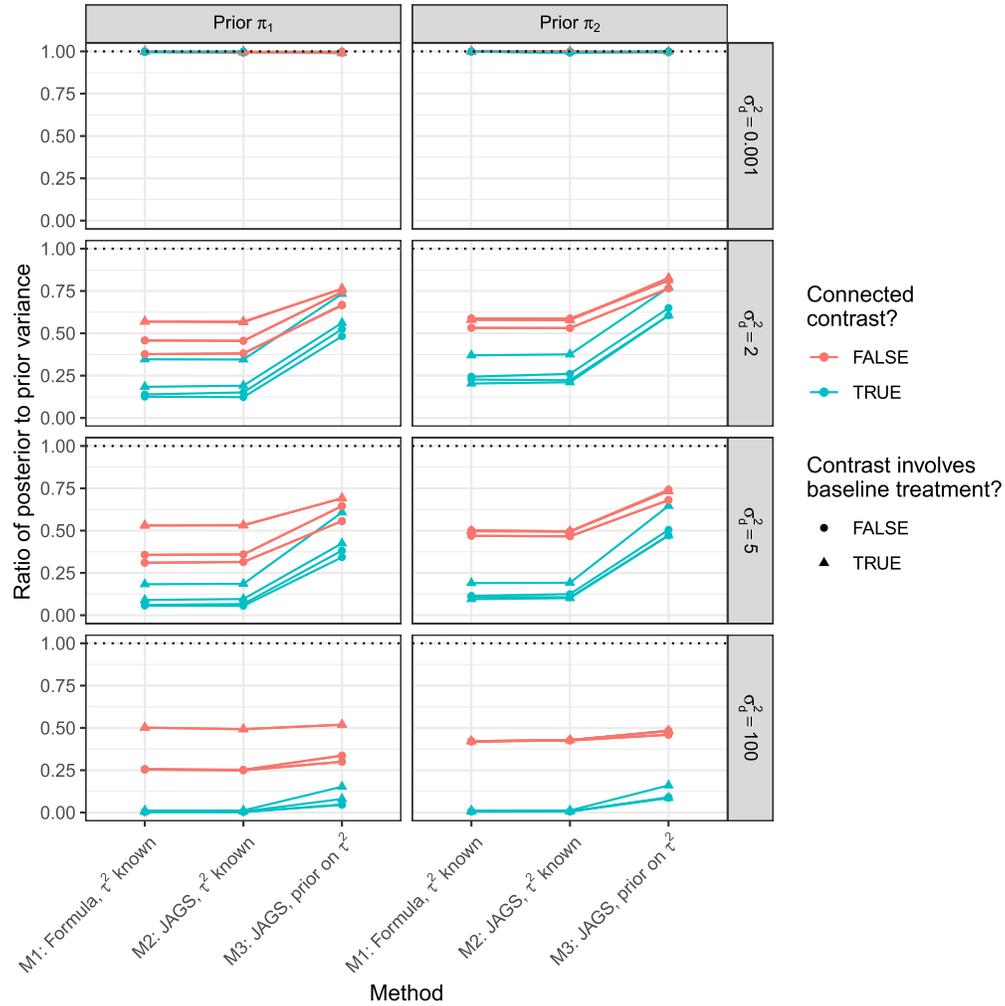
Figure 5: Results for scenario 2, using 15 patients per arm. This graph shows the posterior-to-prior variance ratio for all (10) treatment contrasts obtained by methods M1, M2 and M3, with priors π_1 and π_2 and four different values of σ_d^2. Note that some points and lines appear stacked on top of each other. The dotted horizontal lines indicate the value 1 corresponding to no difference between the posterior and the prior variance.

a logit link. We conduct two analyses: one using prior $\pi_1(\mathbf{d})$ and one using $\pi_2(\mathbf{d})$, both with $\sigma_d^2 = 100$. All other priors are set the same as in Section 5. We choose SK as the reference treatment since it is the most prevalent treatment in the disconnected network (tied with tPA). We use 5 million iterations thinned by a factor of 100, after a burnin period of 100,000 iterations and an adaptation phase of 50,000 iterations (run time 20–25
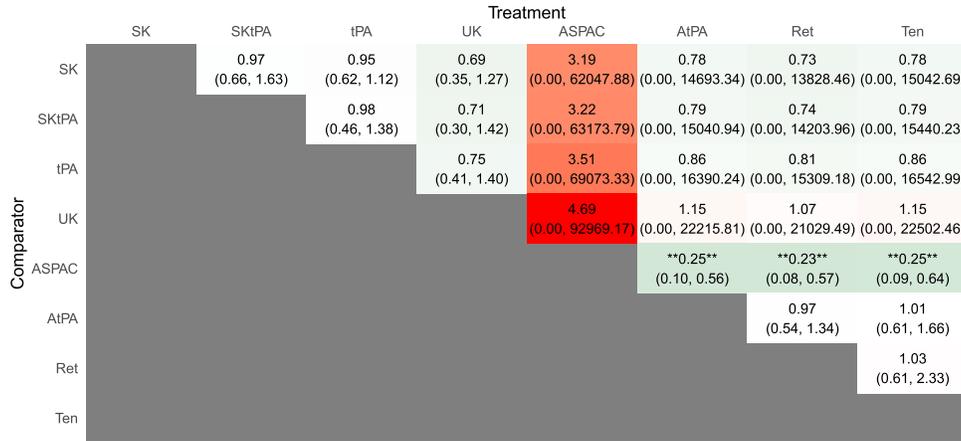
Figure 6: League table representing posterior median odds ratios along with 95% quantile based credible intervals when comparing the treatment at the top with the treatment on the left. Statistical significance is indicated with double asterisk. Red indicates that the estimated odds of mortality is larger for the treatment at the top than the treatment on the left while green indicates a lower estimate.

minutes for 3 chains not in parallel). We used such a large number of iterations because there is a strong autocorrelation in the posterior variance for disconnected contrasts. The R code is provided in Online Appendix.

Figure 6 presents a league table of all estimated pairwise comparisons after conducting the analysis with prior π_1. We find statistical evidence that treatments AtPA, Ret and Ten are more efficacious than ASPAC at preventing mortality. The credible intervals for disconnected contrasts are extremely large, as expected (see upper right quadrant). The same conclusions are observed when conducting the analysis with prior π_2 (league table omitted).

For each contrast, we calculated the ratio of posterior to prior variance of $\mathbf{d}$. When prior π_1 was used, the ratio varied between 0.0002 and 0.0013 for connected contrasts. For disconnected contrasts involving the reference treatment SK, the ratio varied between 0.2563 and 0.2576 which is in line with the lower bound $p^{-1} = 0.25$ calculated using our theoretical results (here we have 4 treatments in the auxiliary subnetwork thus $p = 4$). For other disconnected contrasts, the ratio varied between 0.1282 and 0.1293 which is also in line with the lower bound $0.5p^{-1} = 0.125$ calculated using our theoretical results. When prior π_2 was used, the ratio varied between 0.0002 and 0.0025 for connected contrasts. For disconnected contrasts, the ratio varied between 0.2571 and 0.2588 which is in line with the lower bound $0.5p^{-1}/(1 - pt^{-1}) = 0.25$ calculated using our theoretical results (here we have 8 treatments thus $t = 8$).

An interesting point to note from these results is that the ratios of posterior to prior variance hovered very closely to the theoretical lower bounds.

7   Discussion

In this paper, we investigated theoretically and empirically how evidence flows in disconnected networks when using the contrast-based approach to network meta-analysis. We derived lower bounds for the ratio of posterior to prior variance on disconnected treatment contrasts in the limiting cases of uninformative and point mass priors (Table 1) and illustrated their applicability with the thrombolysis dataset. The lower bounds demonstrate that when using an uninformative prior on the treatment contrasts, the standard NMA approach is not useful to analyze disconnected networks (even when the number of studies, treatments or patients is large). Our theoretical developments assumed that the sample size of the studies was large enough so that the normal approximation to the binomial would hold. Very large or small probabilities of experiencing the outcome in some arms could also affect the validity of the normal approximation. Our theoretical findings thus may not apply in networks where evidence is sparse either because there are only very small studies or because the event is rare or both of these issues. However, the thrombolytic dataset where 29% of arms had no more than 5 events has proven robust to the normality assumption. The simulation study has also proven robust when using 15 patients per arm.

Our simulation study has shown that the standard NMA approach can be useful when using moderately informative priors on $\mathbf{d}$. Moderately informative priors can be used for incorporating information from observational data or non randomized trials in the evidence base for a network meta-analysis (Schmitz et al., 2013; Jenkins et al., 2014; Zhang et al., 2019; Sutton and Abrams, 2001; Mak et al., 2009; Salpeter et al., 2009). In our simulation study, we observed ratios of posterior to prior variance in the range of 0.5 to 0.75 for disconnected contrasts with a sample size of 1000 and 0.55 to 0.8 for a sample size of 15. It was even the case that the ratio of posterior to prior variance for some disconnected contrasts was smaller than that of a connected contrast from the same network, showing that moderately informative priors can really instigate a flow of the evidence in a disconnected network.

The intuitive explanation as to why the contrast-based approach fails under an uninformative prior on the treatment contrasts is that when using independent priors on the α_i's, strength is not borrowed across trials to inform the *absolute* odds of the outcome in each study. One strategy to circumvent this, assuming similarity between trials, is to instead use a common prior on the trial-specific baseline treatment effects (the α_i's) to borrow strength across trials (Goring et al., 2016; Béliveau et al., 2017). This idea is controversial as the α_i's might not be exchangeable when studies are conducted on different patient populations. However, a case study on two datasets showed this strategy to be appropriate for those datasets suggesting that more real-world datasets could benefit from this strategy (Béliveau et al., 2017). Clearly, more theoretical and empirical investigations are required to assess and compare the performance of competing methods for the analysis of disconnected networks. We hope that the theoretical work presented herein and the empirical simulations will inspire such developments.

Supplementary Material

Online Appendix. A Theoretical Investigation on How Evidence Flows in Bayesian Network Meta-Analysis of Disconnected Networks (DOI: 10.1214/20-BA1224SUPPA; .pdf). The Online Appendix includes proofs, Maple code, traceplots from the simulation study and R code for the data analysis.

R Code Supplement (DOI: 10.1214/20-BA1224SUPPB; .zip). The R code Supplement available online includes the R code for the simulation study.

References

Béliveau A., Goring S., Platt W. R. & Gustafson P. (2017). Network meta-analysis of disconnected networks: How dangerous are random baseline treatment effects? *Research Synthesis Methods* **8**, 465–474. 820

Béliveau A., Boyne D. J., Slater J., Brenner D. & Arora P. (2019). BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network Meta-analyses. *BMC Medical Research Methodology* **19**(196). 817

Béliveau A. & Gustafson P. (2020a). Online Appendix. A Theoretical Investigation on How Evidence Flows in Bayesian Network Meta-Analysis of Disconnected Networks. *Bayesian Analysis*. doi: https://doi.org/10.1214/20-BA1224SUPPA. 810

Béliveau A. & Gustafson P. (2020b). R Code Supplement. *Bayesian Analysis*. doi: https://doi.org/10.1214/20-BA1224SUPPB. 814

Boland A., Dundar Y., Bagust A., Haycox A., Hill R., Mota R., Walley T. & Dickson A. (2003). Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation. *Health Technology Assessment* **7**(15), 1–136. MR2831918. doi: https://doi.org/10.1155/2011/523937. 804, 817

Borenstein M., Hedges L. V., Higgins J. P. T. & Rothstein H. R. (2009). *Introduction to Meta-Analysis*. Wiley. 809

Caro J. J. & Ishak, K. J. (2010). No Head-to-Head Trial? Simulate the Missing Arms. *PharmacoEconomics* **28**(10), 957–967. 806

Dias S., Sutton A. J., Ades A. E. & Welton N. J. (2013). Evidence Synthesis for Decision Making 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials. *Medical Decision Making* **33**, 607–617. 804, 805, 806, 807, 808, 809

Dias S. & Ades A. E. (2016). Absolute or Relative Effects? Arm-Based Synthesis of Trial Data. *Research Synthesis Methods* **7**, 23–28. 805

Dias S., Ades A., Welton N. J., Jansen J. P. & Sutton A. J. (2018). *Network meta-analysis for decision-making*. Hoboken: Wiley. 804

Efthimiou O., Debray T. P., van Valkenhoef G., Trelle S., Panayidou K., Moons K. G., Reitsma J. B., Shang A. & Salanti G. (2016). GetReal in network

meta-analysis: a review of the methodology. *Research Synthesis Methods* **7**(3), 236–263.   804

Goring S. M., Gustafson P., Liu Y., Saab S., Cline S. K. & Platt R. W. (2016). Disconnected by design: analytic approach in treatment networks having no common comparator. *Research Synthesis Methods* **7**(4), 420–432.   805, 806, 820

Higgins J. P. T., Thomas J., Chandler J., Cumpston M., Li T., Page M. J. & Welch V. A. (editors) (2019). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)*. Cochrane. Available from www.training.cochrane.org/handbook.   805

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. & 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.   804

Ishak K. J., Proskorovsky I. & Benedict A. (2015). Simulation and Matching-Based Approaches for Indirect Comparison of Treatments. *PharmacoEconomics* **33**(6), 537–549.   806

Jenkins D., Bujkiewicz S., Martina R., Dequen P. & Abrams K. R. (2014). Methods for the inclusion of real world evidence in network meta-analysis. *Value in Health* **17**(7), A576.   820

Lu G. & Ades A. E. (2004). Combination of Direct and Indirect Evidence in Mixed Treatment Comparisons. *Statistics in Medicine* **23**(20), 3105–3124. MR2256166. doi: https://doi.org/10.1198/016214505000001302.   804, 805, 807

Mak A., Cheung M. W. L., Ho R. C. M., Cheak A. A. & Lau C. S. (2009). Bisphosphonates and atrial fibrillation: Bayesian meta-analyses of randomized controlled trials and observational studies. *BMC Musculoskeletal Disorders* **10**, 113.   820

Phillippo D. M., Ades A. E., Dias S., Palmer S., Abrams K. R. & Welton N. J. (2016). *NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE*. Decision Support Unit, 81 p.   806

Phillippo D. M., Ades A. E., Dias, S., Palmer S., Abrams K. R. & Welton N. J. (2018). Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Medical Decision Making* **38**(2), 200–211.   806

Pooley N., Papageorgakopoulou C., Adkins E. & Payne R. (2017). The Use Of Matching Adjusted Indirect Comparison (MAIC) And Simulated Treatment Comparison (STC) In Hta Submissions; Learnings From Recent Submissions. *Value in Health* **20**(9), A769–A770.   806

Rucker G., Petropoulou M. & Schwarzer G. (2019). Network meta-analysis of multicomponent interventions. *Biometrical Journal*, 1–14.   805

Salpeter S. R., Cheng J., Thabane L., Buckley N. S. & Salpeter E. E.

(2009). Bayesian Meta-analysis of Hormone Therapy and Mortality in Younger Post-menopausal Women. *The American Journal of Medecine* **122**, 1016–1022. 820

SIGNOROVITCH J. E., WU E. Q., YU A. P., GERRITS C. M., KANTOR E., BAO Y. ET AL. (2010). Comparative Effectiveness Without Head-to-Head Trials: A Method for Matching-Adjusted Indirect Comparisons Applied to Psoriasis Treatment with Adalimumab or Etanercept. *PharmacoEconomics* **28**(10), 935–945. 806

SIGNOROVITCH J. E., SIKIRICA V., ERDER M. H., XIE J., LU M., HODGKINS P. S., BETTS K. A. & WU E. Q. (2012). Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. *Value in Health* **15**, 940–947. 806

SCHMITZ S., ADAMS R. & WALSH C. (2013). Incorporating data from various trial designs into a mixed treatment comparison model. *Statistics in Medicine* **32**(17), 2935–2949. MR3073827. doi: https://doi.org/10.1002/sim.5764. 820

SCHMITZ S., MAGUIRE A., MORRIS J., RUGGERI K., HALLER E., KUHN I., LEAHY J., HOMER N., KHAN A., BOWDEN J., BUCHANAN V., O'DWYER M., COOK G. & WALSH C. (2018). The use of single armed observational data to closing the gap in otherwise disconnected evidence networks: a network meta-analysis in multiple myeloma. *BMC Medical Research Methodology* **18**(66). 805

STEVENS J. W., FLETCHER C., DOWNEY G. & SUTTON A. (2018). A review of methods for comparing treatments evaluated in studies that form disconnected networks of evidence. *Research Synthesis Methods* **9**(2), 148–162. 805, 806

SUTTON A. J. & ABRAMS K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* **10**, 277–303. 820

THOM H., WHITE I. R., WELTON N. J. & LU G. (2018). Automated methods to test connectedness and quantify indirectness of evidence in network meta-analysis. *Research Synthesis Methods* **10**, 113–124. 805

ZARIN W., VERONIKI A. A., NINCIC V., VAFAEI A., REYNEN E., MOTIWALA S. S. ET AL. (2017). Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Medicine* **15**(3), 1–11. 803

ZHANG K., ARORA P., SATI N., BÉLIVEAU A., TROKE N., VERONIKI A. A., RODRIGUES M., RIOS P., ZARIN W. & TRICCO A. C. (2019). Characteristics and methods of incorporating randomized and non-randomized evidence in network meta-analyses: a scoping review *Journal of Clinical Epidemiology* **113**, 1–10. 820