Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2021, Vol. 30(1) 151–165

DOI: 10.1177/0962280220945731
journals.sagepub.com/home/smm
SAGE

A variance shrinkage method improves arm-based Bayesian network meta-analysis

Zhenxun Wang[1], Lifeng Lin[2], James S Hodges[1], Richard MacLehose[3] and Haitao Chu[1]

Abstract

Network meta-analysis is a commonly used tool to combine direct and indirect evidence in systematic reviews of multiple treatments to improve estimation compared to traditional pairwise meta-analysis. Unlike the contrast-based network meta-analysis approach, which focuses on estimating relative effects such as odds ratios, the arm-based network meta-analysis approach can estimate absolute risks and other effects, which are arguably more informative in medicine and public health. However, the number of clinical studies involving each treatment is often small in a network meta-analysis, leading to unstable treatment-specific variance estimates in the arm-based network meta-analysis approach when using non- or weakly informative priors under an unequal variance assumption. Additional assumptions, such as equal (i.e. homogeneous) variances for all treatments, may be used to remedy this problem, but such assumptions may be inappropriately strong. This article introduces a variance shrinkage method for an arm-based network meta-analysis. Specifically, we assume different treatment variances share a common prior with unknown hyperparameters. This assumption is weaker than the homogeneous variance assumption and improves estimation by shrinking the variances in a data-dependent way. We illustrate the advantages of the variance shrinkage method by reanalyzing a network meta-analysis of organized inpatient care interventions for stroke. Finally, comprehensive simulations investigate the impact of different variance assumptions on statistical inference, and simulation results show that the variance shrinkage method provides better estimation for log odds ratios and absolute risks.

Keywords

Bayesian inference, variance prior, network meta-analysis, variance shrinkage method

1 Introduction

Evidence-based practice (EBP) is a powerful theoretical framework to connect study findings to a profession's body of knowledge.[1] Evaluating evidence needed for EBP or scientific research is generally more complicated. Typically, a hierarchy is used to rank order the available evidence based on quality, with systematic reviews and meta-analyses ranking at the top. In public health, systematic reviews help researchers and practitioners remain up to date with accumulating evidence and identify topics for which further scientific studies are needed. Meta-analysis, on the other hand, by aggregating effects, can address certain biases (e.g. reporting bias and small-study effects) of estimated treatment effects.[2] Network meta-analysis (NMA) was developed to simultaneously compare multiple (more than two) interventions. Compared to pairwise comparison from a traditional meta-analysis,

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA
[2]Department of Statistics, Florida State University, Tallahassee, FL, USA
[3]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA

Corresponding author:
Haitao Chu, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA.
Email: chux0051@umn.edu

NMA can gain precision by considering both direct and indirect comparisons and has the potential to rank regimens more explicitly.[3] This article concentrates on Bayesian hierarchical models[4,5] to perform NMAs, which are widely applied in practice.[6,7]

Two broadly used Bayesian hierarchical approaches have been considered, i.e. the contrast-based (CB) approach[4,8,9] and the arm-based (AB) approach.[5,10,11] AB-NMA focuses on absolute treatment effects and assumes that absolute effects are exchangeable across studies, while the CB method assumes that relative effects (contrasts) are exchangeable across trials, a difference that has led, among other things, to a debate over random baseline treatment effects.[12–14] However, the primary advantage of AB-NMA is that it naturally estimates absolute risks (ARs) and AR differences. ARs are essential to calculate the incremental cost-effectiveness ratio, a useful decision tool for resource allocation. Also, results from AB-NMA are less sensitive to treatment exclusions.[15] The main challenge in NMA is a lack of information because a treatment was included in a limited number of studies, and few arms are included in most trials. Specifically, in a randomized clinical trial, healthcare practitioners commonly select only two to four regimens (in most cases, only two) from a list of treatments, based on previous findings and published results. This creates two main obstacles in analyzing NMAs. First, not all treatments are directly compared in an NMA; in an empirical study, 18.8% of NMAs are "star-shaped," i.e. all active treatments were compared in trials only to a control.[16] This phenomenon may produce difficulties in accurately estimating correlations among treatments in the AB approach.[17] In the CB approach, it may cause variances of certain contrasts to be overestimated.[9] Second, the number of clinical studies involving each treatment is limited. For example, we extracted 42 NMAs with binary outcomes from a total of 186 NMAs investigated by Nikolakopoulou et al.[16] Descriptive statistics of these 42 networks (Figure 1) show that nearly 40% of treatments in these NMAs are included in four or fewer clinical studies, which can lead to overestimation of variances of relevant treatment effects in the AB approach if noninformative priors are used.[17] To overcome this problem, both the AB and CB approaches often make additional assumptions. For example, in the CB model, Dias et al.[18] advocated assuming homogeneous between-study variances, that is, all treatment contrasts are assumed to have a common between-study variance. Similarly, in the AB model, we may assume that all treatments share the same between-study variance.

However, a homogeneous treatment-specific variance assumption may not be valid in the AB approach. Motivated by the James–Stein estimator[19] and the double shrinkage estimator,[20–22] this article proposes a new method to relax this potentially strong assumption. While the James–Stein and double shrinkage estimators can only be applied to the classical normal mean problem with "known and equal" variance and "unknown and unequal" variances, respectively, our method can be applied to multivariate normal problems with a focus on variance shrinkage. By assuming that different treatment variances share a common prior instead of assuming

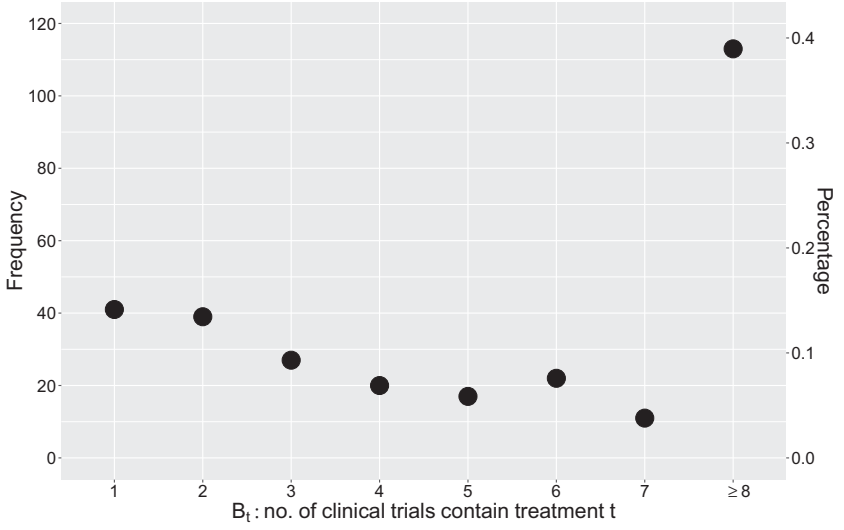

Figure 1. Dot plot describing 42 NMAs with binary outcomes, from a total of 186 NMAs investigated by Nikolakopoulou et al.[16] The *x*-axis denotes B_t: the number of clinical trials containing a certain treatment *t*. The *y*-axis is the frequency and percentage of such treatments in each category. Nearly 40% of treatments in these NMAs are included in four or fewer clinical trials.

they are simply equal, we not only use a less strict assumption but also can estimate variances in a data-dependent way when insufficient data are available.

The rest of this article is organized as follows. The next section describes a motivating example of an NMA of organized inpatient care for stroke. This is followed by a brief review of the AB model for analyzing NMA datasets with dichotomous outcomes and introduces the variance shrinkage method. Then, the results from applying the variance shrinkage method to the motivating example are presented, followed by extensive simulation studies in the penultimate section, comparing the performance of different priors on variances. The final section discusses our findings and directions for future research.

2 Motivating example and notation

A stroke occurs when oxygen-rich blood flow to the brain is blocked, which leads to brain cell death. It is currently the world's second leading cause of mortality[23] and the third leading cause of disability.[24] As of the early 2000s, there were debates about whether organized inpatient (stoke unit) care, a multidisciplinary team specializing in stroke management, could increase patient survival and recovery.[25] Five types of organized inpatient care had been examined: (1) stroke ward, (2) general medical ward, (3) mixed rehabilitation ward, (4) mobile stroke team, and (5) acute (semi-intensive) ward. The Stroke Unit Trialists' Collaboration[26] carried out a systematic review on organized inpatient (stroke unit) care for stroke, including 28 studies with 6585 participants. The outcome was death by the end of scheduled follow-up. Figure 2 is a network plot of the studies with the five treatments.

To better understand the problems motivating the present work, we first specify basic notation for an NMA with binary outcomes. Assume an NMA includes K studies (e.g. $K=28$ in this case) comparing a total of T treatments (e.g. $T=5$). No study includes all T treatments; each includes only a subset of the treatments. In particular, A_k ($k = 1, \ldots, K$) denotes the subset of treatments in the k^{th} study; for example, $A_4 = \{1, 2, 3\}$ implies that treatments 1, 2, and 3 are compared in the fourth study. Generally, the number of elements in the set A_k (denoted by $|A_k|$) is between 2 and 4, as few clinical studies compare more than 4 treatments at the same time. We further define the number of studies containing the t^{th} treatment as B_t, and the number of direct comparisons between treatments i and j as C_{ij}. Let $D = \{D_1, \ldots, D_K\}$ be the data collected with D_k representing the data from the k^{th} study. Then, for NMAs with dichotomous outcomes, $D_k = \{(r_{kt}, n_{kt}), t \in A_k\}$, with r_{kt} and n_{kt} denoting the numbers of events and participants in the t^{th} treatment group in the k^{th} study, respectively.

In this motivating example, $B_5 = 2$ and some C_{ij}'s are ≤ 2, which may be considered as examples of the "lack of information" situation described in Section 1. We reanalyzed the stoke data using the AB-NMA approach specified by Zhang et al.[5] Specifically, we used the exchangeable correlation structure with a uniform prior $U(-\frac{1}{T-1}, 1)$ on the correlation coefficients[27] and the heterogeneous variance assumption with separate uniform priors $U(0, 5)$ on the standard deviations (henceforth referred to as the UV approach). Figure 3(a) and (b) presents

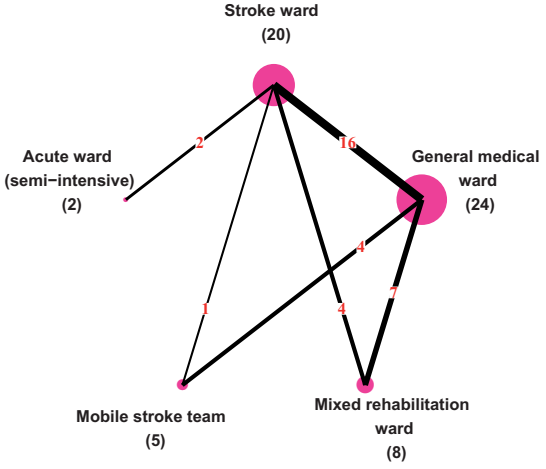

Figure 2. Network plot of the case study of organized inpatient care for stroke. Each node in the plot represents a treatment, and each edge represents a direct comparison between two treatments. Vertex radius is proportional to B_t (the number of studies containing treatment t), and the edge thickness is proportional to C_{ij} (the number of direct comparisons between treatments i and j).

the forest plots of the standard deviations and ARs of treatments 1 to 5. The UV method's results are olive-colored; the other results will be explained later. Clearly, using the UV approach, the posterior distribution of the standard deviation of acute (semi-intensive) ward, for which the 95% credible interval (CrI) was (0.07, 4.20), is dominated by the $U(0, 5)$ prior distribution. Similarly, for acute (semi-intensive) ward, the 95% CrI for the risk was extremely wide. To overcome these problems, we could use the homogeneous variance assumption instead and place a $U(0, 5)$ prior on the common standard deviation (henceforth referred to as the EV approach). However, this strong assumption forces the standard deviations of mobile stroke team and mixed rehabilitation ward to take a common value, which may not be tenable according to the UV method's estimate. To achieve a trade-off between the UV and EV approaches, the following section proposes a variance shrinkage method.

3 Methods

3.1 AB Bayesian NMA

This subsection gives a brief introduction to AB-NMA.[5] Generally speaking, the AB-NMA model has two levels. The first level is within study, at which NMAs with different types of outcomes would have different models. The second level is between study, where all studies share a distribution with different link functions. We focus on NMA with binary outcomes here. The underlying model is

$$\begin{aligned} &\text{Level I}: r_{kt} \sim \text{Binomial}(n_{kt}, p_{kt}),\ t \in A_k,\ k = 1, \ldots, K; \\ &\text{Level II}: \text{logit}(p_{kt}) = \theta_{kt}; \quad (\theta_{k1}, \ldots, \theta_{kT})' \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{aligned} \tag{1}$$

where p_{kt} is the probability of an event (i.e. AR) for the t^{th} treatment in the k^{th} study, and the vector $\boldsymbol{\theta}_k = (\theta_{k1}, \ldots, \theta_{kT})'$ follows the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Here, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)'$ contains the overall logit event rate (i.e. log odds) for each treatment, and $\boldsymbol{x}'$ denotes the transpose of the vector $\boldsymbol{x}$. If we denote the between-study standard deviation for treatment t by δ_t, we can decompose $\boldsymbol{\Sigma}$ as $\boldsymbol{\Delta}\mathbf{P}\boldsymbol{\Delta}$, where $\mathbf{P} = \{\rho_{ij}\}$ is the correlation matrix, and $\boldsymbol{\Delta}$ is a diagonal matrix with δ_t being its t^{th} diagonal element.

3.2 Prior specifications

Prior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ need to be specified. We set weakly informative priors $N(0, 100^2)$ on μ_t $(t = 1, \ldots, T)$. For the covariance matrix Σ, we use the separation strategy proposed by Barnard et al.[28]

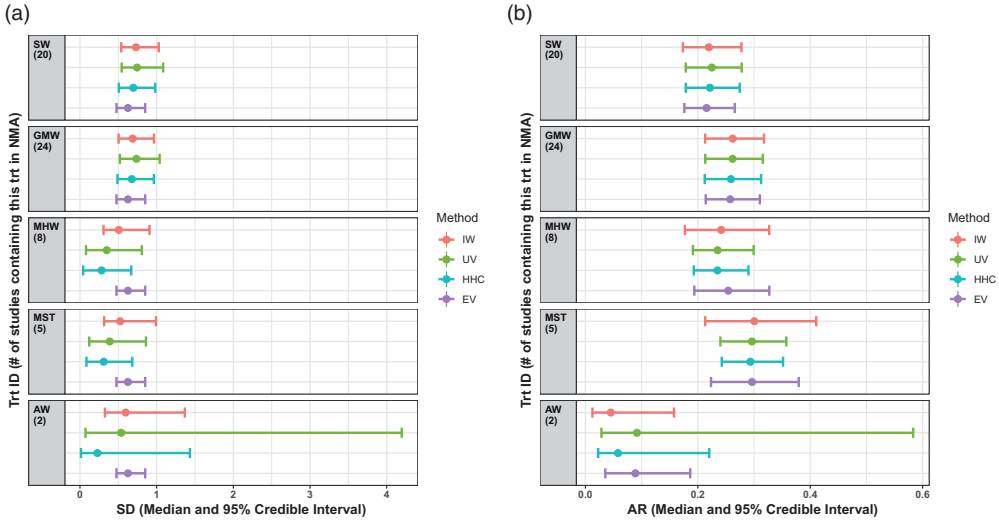

Figure 3. Results for case study of organized inpatient care for stroke: forest plot of (a) standard deviations δ_t and (b) absolute risk p_t (posterior median with 95% credible interval). Different colors indicate different priors. The *y*-axis represents the treatment label, with B_t in parentheses. Treatment labels: (1) stroke ward, (2) general medical ward, (3) mixed rehabilitation ward, (4) mobile stroke team, and (5) acute (semi-intensive) ward.
IW: inverse-Wishart; HHC: hierarchical half-Cauchy; NMA: network meta-analysis.

Instead of treating the covariance matrix as a whole, this method first decomposes it into separate parts as $\mathbf{\Sigma} = \mathbf{\Delta P \Delta}$ and then sets priors independently on the correlation matrix $\mathbf{P}$ and the standard deviations δ_t $(t = 1, \ldots, T)$, which form the diagonal matrix $\mathbf{\Delta}$. Here, we will simply use the exchangeable correlation structure to set a prior for the correlation matrix $\mathbf{P}$[17,29]; that is, all correlation coefficients ρ_{ij} are assumed equal to ρ, and the uniform prior $U(-\frac{1}{T-1}, 1)$ is assigned to ρ. The lower bound of this uniform prior guarantees that the correlation matrix is positive definite.

3.3 Variance shrinkage method

As mentioned in Section 2, our goal is to achieve a trade-off between the homogeneous and heterogeneous variance assumptions. For this purpose, we propose a less stringent assumption: all δ_ts follow the same prior density with some unknown hyperparameters, and we use the data to estimate the hyperparameters.

Accordingly, we propose the hierarchical half-Cauchy (HHC) prior, denoted by $\text{HHC}(\epsilon_l, \epsilon_u)$, on δ_t $(t = 1, \ldots, T)$; that is, δ_t is distributed as half-Cauchy (HC) with hyperparameter a, which has a uniform prior $U(\epsilon_l, \epsilon_u)$. The HC prior is commonly used for standard deviations[30] and has density $\text{HC}(a) \propto (1 + \delta_t^2/a^2)^{-1}$. Like the uniform prior, the HC prior may also result in overestimation of variances in an AB-NMA when the scale parameter a is large (e.g. $a = 5$ is a common choice). On the other hand, if a is too small, the HC distribution is no longer weakly informative because it has high density near zero (e.g. $a = 0.5$ as in Figure 4). By using the HHC prior, the data-driven posterior distribution of the hyperparameter a determines how informative the prior on δ_t should be: if it is less informative (e.g. $a = 5$ as in Figure 4), it shrinks δ_t less; if it is more informative (e.g. $a = 0.5$ as in Figure 4), it shrinks δ_t more.

3.4 Likelihood and posterior estimation

The likelihood function for $\boldsymbol{\theta}_k$ based on data D_k from the k^{th} study can be written as

$$L(\boldsymbol{\theta}_k | D_k) = \prod_{t \in A_k} \left[\text{logit}^{-1}(\theta_{kt})\right]^{r_{kt}} \left[1 - \text{logit}^{-1}(\theta_{kt})\right]^{n_{kt} - r_{kt}} \tag{2}$$

Denote the aforementioned prior distributions for μ_t, δ_t, a and ρ by $\pi(\mu_t)$, $\pi(\delta_t|a)$, $\pi(a)$ and $\pi(\rho)$, respectively. If the density function of the multivariate normal distribution is $p(\boldsymbol{\theta}_k | \boldsymbol{\mu}, \mathbf{\Sigma}) = p(\boldsymbol{\theta}_k | \boldsymbol{\mu}, \mathbf{\Delta}, \rho)$, the joint posterior

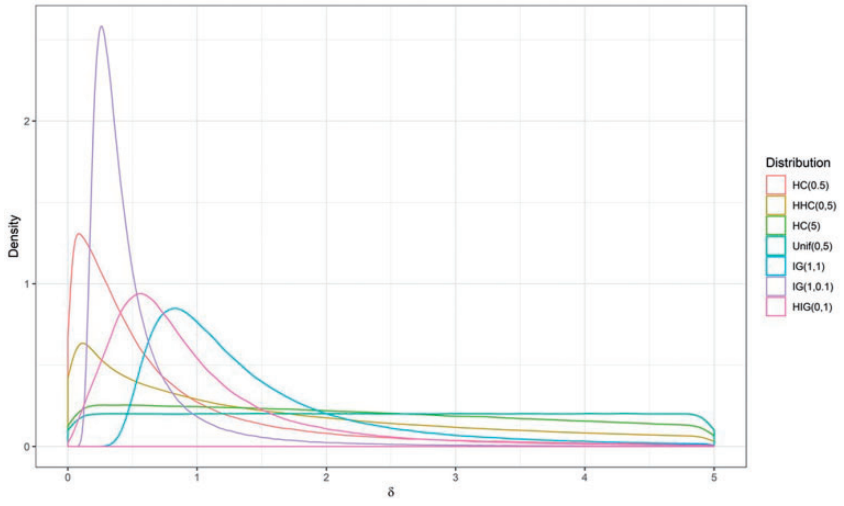

Figure 4. Densities of different priors on standard deviation δ_t. For better visualization, the horizontal axis is limited to $[0, 5]$. HC: half-Cauchy; HHC: hierarchical half-Cauchy; HIG: hierarchical inverse-gamma.

distribution is

$$\begin{aligned}\pi(\mu, \boldsymbol{\Delta}, a, \rho, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K | D) \propto \\ \times \prod_{k=1}^{K} \left\{ \prod_{t \in A_k} \left[\text{logit}^{-1}(\theta_{kt})\right]^{r_{kt}} \left[1 - \text{logit}^{-1}(\theta_{kt})\right]^{n_{kt}-r_{kt}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\theta}_k - \mu)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_k - \mu)} \right\} \\ \times \prod_{t=1}^{T} \pi(\mu_t) \pi(\delta_t | a) \pi(a) \pi(\rho)\end{aligned} \tag{3}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. We use Markov chain Monte Carlo (MCMC) to sample from the joint posterior distribution. The marginal event rate of treatment t is $p_t = E[p_{kt} | \mu_t, \delta_t]$; for the logit link used in equation (1), p_t can be approximated by Zeger et al.[31]

$$\left[1 + \exp\left(-\mu_t \Big/ \sqrt{1 + \frac{256}{75\pi^2}\delta_t^2}\right)\right]^{-1}$$

Two log odds ratio estimands can be considered: (1) the marginal log odds ratio between treatments i and j, $\text{mLOR}_{ij} = \log\left(\frac{p_i/(1-p_i)}{p_j/(1-p_j)}\right)$ and (2) the conditional log odds ratio $\text{cLOR}_{ij} = \mu_i - \mu_j$, which is more common in the meta-analyses literature. In each MCMC iteration, draws of p_t, mLOR_{ij}, and cLOR_{ij} can be calculated using the above equations. Finally, we can make statistical inferences using posterior medians, means, and 95% equal-tailed CrIs estimated from these posterior samples.

4 Data analysis: Organized inpatient care for stroke

This section applies the variance shrinkage method to the motivating example and compares its results with those of other common priors. Specifically, we consider the following models:

- Model 1: The inverse-Wishart (IW) prior, the conjugate prior for the multivariate normal. The prior for the covariance matrix $\boldsymbol{\Sigma}$ is $IW_T(\mathbf{I}, T+1)$, where $T+1$ is the degrees of freedom, and the scale matrix is the $T \times T$ identity matrix $\mathbf{I}$.
- Model 2: The heterogeneous variance assumption (UV). We use the separation strategy with equal correlations (all $\rho_{ij} = \rho$) but unequal variances, and put the priors $U\left(-\frac{1}{T-1}, 1\right)$ on ρ and $U(0, 5)$ on each δ_t.
- Model 3: The variance shrinkage method (HHC). It shares the same setting with Model 2 except that the HHC prior $\text{HHC}(0, 5)$ is used for δ_t.
- Model 4: The homogeneous variance assumption (EV). We use the separation strategy with equal correlations and equal variances (all $\delta_t = \delta$). Similarly, we put $U(-\frac{1}{T-1}, 1)$ on ρ and $U(0, 5)$ on δ.

Model 3 "splits the difference" between Models 2 and 4. All models use the vague prior $N(0, 100^2)$ on μ_t ($t = 1, \ldots, T$). We use posterior medians and 95% equal-tailed CrIs as point and interval estimates, respectively.

The Bayesian approach has the advantage of conveniently allowing inferences about treatment rankings. We use the surface under the cumulative ranking (SUCRA) proposed by Salanti et al.[32] as a measure for comparing treatments. Specifically, let prob_{ti} be the probability that treatment t has the i^{th} rank, where $i = 1$ represents the best treatment. The SUCRA of the t^{th} treatment can be calculated as:

$$\text{SUCRA}_t = \frac{1}{T-1} \sum_{j=1}^{T-1} \sum_{i=1}^{j} \text{prob}_{ti}.$$

4.1 Model comparison

We evaluated the models using the deviance information criterion (DIC) by Spiegelhalter et al.[33] and the widely applicable information criteria (WAIC).[34] DIC is the sum of the mean deviance $\overline{D}$ (describing the goodness of fit)

and the effective number of parameters p_D (penalizing for model complexity). A difference larger than 5 in DIC may indicate that the model with lower DIC gives a considerable improvement.[35] Specifically, DIC is calculated as

$$\mathrm{DIC} = \overline{D} + p_D \tag{4}$$

where $\overline{D} = \sum_{k=1}^{K} \sum_{t \in A_k} \overline{\mathrm{Dev}}_{kt}$ and $p_D = \sum_{k=1}^{K} \sum_{t \in A_k} \left(\overline{\mathrm{Dev}}_{kt} - \widetilde{\mathrm{Dev}}_{kt}\right)$. For an NMA model, Dev_{kt} represents the residual deviance for treatment t in study k:

$$\mathrm{Dev}_{kt} = 2\left\{ r_{kt} \log\left(\frac{r_{kt}}{\hat{r}_{kt}}\right) + (n_{kt} - r_{kt}) \log\left(\frac{n_{kt} - r_{kt}}{n_{kt} - \hat{r}_{kt}}\right) \right\}, \quad t \in A_k,\ k = 1, \ldots, K.$$

where $\hat{r}_{kt} = n_{kt} p_{kt}$ is the expected event count of the t^{th} treatment in the k^{th} study. Then, $\overline{\mathrm{Dev}}_{kt}$ is the posterior mean of Dev_{kt}, and $\widetilde{\mathrm{Dev}}_{kt}$ is the residual deviance evaluated at the posterior mean event count $\tilde{r}_{kt} = n_{kt} \overline{p}_{kt}$:

$$\widetilde{\mathrm{Dev}}_{kt} = 2\left\{ r_{kt} \log\left(\frac{r_{kt}}{\tilde{r}_{kt}}\right) + (n_{kt} - r_{kt}) \log\left(\frac{n_{kt} - r_{kt}}{n_{kt} - \tilde{r}_{kt}}\right) \right\}.$$

The other criterion, WAIC, is asymptotically equivalent to leave-one-out cross-validation.[34] Compared to DIC, WAIC is more relevant in a predictive context, because it is calculated as the sum of the posterior distribution (i.e. log pointwise predictive density (lppd)) with a bias correction p_W (i.e. the sum of the variance of individual terms in the log predictive density):

$$\begin{aligned} \mathrm{WAIC} &= -2\overline{\mathrm{lppd}} + 2\overline{p_W}; \\ \mathrm{lppd} &= \sum_{k=1}^{K} \sum_{t \in A_k} \log \int p(r_{kt}, n_{kt} | \boldsymbol{\psi}) \pi(\boldsymbol{\psi} | D)\, \mathrm{d}\boldsymbol{\psi}; \\ p_W &= \sum_{k=1}^{K} \sum_{t \in A_k} \mathrm{Var}(\log(p(r_{kt}, n_{kt} | \boldsymbol{\psi}))) \end{aligned} \tag{5}$$

where $\pi(\boldsymbol{\psi}|D)$ is the joint posterior distribution in equation (3), and $\boldsymbol{\psi}$ is all unknown parameters including $\boldsymbol{\mu}$, $\boldsymbol{\Delta}$, a, ρ, and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$. Let $\{\boldsymbol{\psi}^s, s = 1, \ldots, S\}$ be MCMC samples of $\boldsymbol{\psi}$ from this joint posterior distribution, then lppd and p_W can be estimated as

$$\begin{aligned} \overline{\mathrm{lppd}} &= \sum_{k=1}^{K} \sum_{t \in A_k} \log\left(\frac{1}{S} \sum_{s=1}^{S} p(r_{kt}, n_{kt} | \boldsymbol{\psi}^s)\right); \\ \overline{p_W} &= \sum_{k=1}^{K} \sum_{t \in A_k} \frac{1}{S-1} \sum_{s=1}^{S} \left(\log(p(r_{kt}, n_{kt} | \boldsymbol{\psi}^s)) - \overline{\log(p(r_{kt}, n_{kt} | \boldsymbol{\psi}^s))}\right)^2 \end{aligned} \tag{6}$$

For both DIC and WAIC, a model with a smaller value is favored.

4.2 Results

Online Appendix A provides the diagnostic plots of HHC method (Model 3). Based on trace plots and autocorrelation plots of δ_t, p_t, and mLOR_{ij}, the MCMC samples have converged well.

Table 1 presents results for the marginal log odds ratios mLOR_{ij}, the AR of treatment t (p_t), the standard deviation of treatment t's effect (δ_t), the SUCRA of treatment t (a smaller value indicates worse performance in preventing death), and DIC. The IW prior had worse performance than the other three priors in terms of DIC, as the differences in DIC were larger than 5 relative to other models. The EV prior was worse than the HHC, UV, and IW priors in terms of goodness of fit, as its mean deviance $\overline{D}$ was highest with 58.68. This suggests that the homogeneous variance assumption for the EV prior is questionable in this NMA. In addition, WAIC provided

Table 1. Organized inpatient care for stroke data: comparing posterior median and 95% credible intervals under four models (IW, UV, HHC, and EV); $mLOR_{ij}$ compares the i^{th} and j^{th} treatment, absolute risk of events for the t^{th} treatment (p_t), standard deviation of the t^{th} treatment (δ_t), and SUCRA of the t^{th} treatment ($SUCRA_t$).

	Point estimate (95% credible interval)			
Parameter	IW	UV	HHC	EV
$mLOR_{12}$	–0.23 (–0.50, 0.03)	–0.19 (–0.39, –0.03)	–0.20 (–0.38, –0.03)	–0.23 (–0.41, –0.06)
$mLOR_{13}$	–0.13 (–0.59, 0.34)	–0.07 (–0.41, 0.26)	–0.08 (–0.39, 0.24)	–0.21 (–0.52, 0.09)
$mLOR_{14}$	–0.42 (–0.97, 0.10)	–0.36 (–0.69, –0.06)	–0.38 (–0.68, –0.07)	–0.42 (–0.76, –0.08)
$mLOR_{15}$	1.78 (0.39, 3.16)	1.03 (–1.56, 2.31)	1.53 (0.03, 2.53)	1.03 (0.20, 2.00)
$mLOR_{23}$	0.10 (–0.34, 0.56)	0.13 (–0.19, 0.42)	0.12 (–0.17, 0.43)	0.02 (–0.29, 0.32)
$mLOR_{24}$	–0.19 (–0.72, 0.31)	–0.18 (–0.46, 0.12)	–0.17 (–0.46, 0.12)	–0.19 (–0.51, 0.14)
$mLOR_{25}$	2.01 (0.62, 3.38)	1.24 (–1.36, 2.51)	1.73 (0.23, 2.73)	1.27 (0.42, 2.24)
$mLOR_{34}$	–0.30 (–0.91, 0.32)	–0.30 (–0.66, 0.10)	–0.30 (–0.64, 0.05)	–0.21 (–0.64, 0.22)
$mLOR_{35}$	1.91 (0.48, 3.30)	1.12 (–1.49, 2.40)	1.61 (0.09, 2.62)	1.25 (0.37, 2.25)
$mLOR_{45}$	2.20 (0.75, 3.62)	1.41 (–1.21, 2.69)	1.91 (0.38, 2.92)	1.46 (0.57, 2.48)
p_1	0.22 (0.17, 0.28)	0.22 (0.18, 0.28)	0.22 (0.18, 0.27)	0.22 (0.18, 0.27)
p_2	0.26 (0.21, 0.32)	0.26 (0.21, 0.32)	0.26 (0.21, 0.31)	0.26 (0.21, 0.31)
p_3	0.24 (0.18, 0.33)	0.24 (0.19, 0.30)	0.23 (0.19, 0.29)	0.25 (0.19, 0.33)
p_4	0.30 (0.21, 0.41)	0.30 (0.24, 0.36)	0.29 (0.24, 0.35)	0.30 (0.22, 0.38)
p_5	0.05 (0.01, 0.16)	0.09 (0.03, 0.58)	0.06 (0.02, 0.22)	0.09 (0.04, 0.19)
δ (equal variance)	.	.	.	0.63 (0.48, 0.85)
δ_1	0.73 (0.54, 1.03)	0.74 (0.54, 1.09)	0.70 (0.51, 0.98)	.
δ_2	0.69 (0.50, 0.97)	0.74 (0.52, 1.04)	0.68 (0.49, 0.96)	.
δ_3	0.51 (0.31, 0.91)	0.35 (0.08, 0.81)	0.28 (0.04, 0.67)	.
δ_4	0.52 (0.32, 0.99)	0.39 (0.12, 0.86)	0.31 (0.08, 0.68)	.
δ_5	0.60 (0.33, 1.37)	0.54 (0.07, 4.20)	0.23 (0.01, 1.43)	.
$SUCRA_1$(%)	65	71	67	73
$SUCRA_2$(%)	28	29	28	33
$SUCRA_3$(%)	46	51	52	37
$SUCRA_4$(%)	11	12	05	09
$SUCRA_5$(%)	99	86	98	100
DIC	99.99	93.66	90.27	94.53
$\overline{D}$	53.93	56.00	54.50	58.68
p_D	46.05	37.66	35.77	35.86
WAIC	6742.95	6741.44	6738.99	6747.61

Treatment labels: (1) stroke ward, (2) general medical ward, (3) mixed rehabilitation ward, (4) mobile stroke team, and (5) acute (semi-intensive) ward. IW: inverse-Wishart; HHC: hierarchical half-Cauchy; CP: coverage probability; SUCRA: surface under the cumulative ranking; WAIC: widely applicable information criteria; DIC: deviance information criterion.

similar results as DIC, with EV method performed worst; WAIC for EV, UV, and HHC was 6747.61, 6741.44, and 6738.99, respectively.

Figure 3(a) is a forest plot of the standard deviations δ_t. All priors gave almost the same results for δ_1 and δ_2, because sufficient information was available for these two treatments ($B_1 = 20$ and $B_2 = 24$). The HHC prior gave results more similar to the UV prior than to the EV prior for δ_3 and δ_4; the posterior median and 95% CrI of δ_3 were 0.28 (0.04, 0.67) for the HHC prior, 0.35 (0.08, 0.81) for the UV prior, and 0.63 (0.48, 0.85) for the EV prior; the results were quite similar for δ_4. As $B_3 = 8$ and $B_4 = 5$, the information available for treatments 3 and 4 might be sufficient, and we might be more confident in the results given by the UV prior than those given by the EV prior. In particular, the EV prior might overestimate δ_3 and δ_4 due to the strong assumption of equal standard deviations. For δ_5, the UV model gave an extremely wide interval, not surprising given that $B_5 = 2$, while the EV model gave an interval much narrower than either IW or HHC, with the latter splitting the difference between UV and EV.

The difference in variance estimates affects the estimates of ARs, as shown in Figure 3(b). Specifically, the EV prior yielded a wider CrI for mixed rehabilitation ward; the posterior median and 95% CrI were 0.23 (0.19, 0.29) using the HHC prior, 0.24 (0.19, 0.30) using the UV prior, and 0.25 (0.19, 0.33) using the EV prior, while for mobile stroke team, these were 0.29 (0.24, 0.35) for the HHC prior, 0.30 (0.24, 0.36) for the UV prior, and 0.30

(0.22, 0.38) for the EV prior. On the other hand, the UV prior produced a wide CrI for p_5 because the information was limited for acute (semi-intensive) ward ($B_5 = 2$), and the posterior distribution was thus greatly influenced by prior information. The HHC prior could borrow some information about the standard deviation of acute (semi-intensive) ward from other δ_t, which yielded a much narrower CrI for the AR; specifically, the 95% CrI lengths were 0.20, 0.55, and 0.15 using the HHC, UV, and EV priors, respectively.

Estimated log odds ratios and SUCRAs were also similar for the UV and HHC priors for mixed rehabilitation ward and mobile stroke team, and for the HHC and EV priors for acute (semi-intensive) ward. In particular, acute (semi-intensive) ward was very likely the best treatment based on the EV and HHC priors, while using the UV prior, the performance of acute (semi-intensive) ward had large uncertainty (SUCRAs were 0.86, 0.98, and 1.00 using the UV, HHC, and EV priors, respectively). SUCRAs indicated that mixed rehabilitation ward was the third best treatment under the HHC and UV priors, while the SUCRA for mixed rehabilitation ward (0.37) was close to that for general medical ward (0.33) under the EV prior. For the IW prior, we could not claim that stroke ward was significantly better than general medical ward; the mLORs with 95% CrIs were –0.23 (–0.50, 0.03) for the IW prior, –0.19 (–0.39, –0.03) for the UV prior, –0.20 (–0.38, –0.03) for the HHC prior, and –0.23 (–0.41, –0.06) for the EV prior. For comparing mobile stroke team versus stroke ward, the mLORs with 95% CrIs were –0.42 (–0.97, 0.10) for the IW prior, –0.36 (–0.69, –0.06) for the UV prior, –0.38 (–0.68, –0.07) for the HHC prior, and –0.42 (–0.76, –0.08) for the EV prior; these relative effects were significant under all priors except the IW.

In summary, when the homogeneous variance assumption may not be valid, the HHC prior can provide more reasonable results than the EV prior. At the same time, unlike the UV prior, the HHC prior allows treatments with limited data to borrow information from other treatments to estimate parameters.

5 Simulation studies

5.1 Simulation settings

We conducted comprehensive simulation studies to compare the four priors defined and used in Section 4. Each simulated NMA dataset had $K = 20$ studies and $T = 6$ treatments (denoted 1 to 6). The number of participants in each treatment arm in each study, n_{kt}, was fixed at 200. The number of simulated datasets in each simulation setting was 1000.

We generated a complete dataset under the AB model with binary outcomes as in equation (1) with $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6)' = (-2, -2.5, -3, -2, -1.5, -3)'$ and $(\theta_{k1}, \ldots, \theta_{k6})' \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{\Delta}\mathbf{P}\boldsymbol{\Delta}$. The correlation matrix $\mathbf{P}$ had an exchangeable structure with all off-diagonal entries 0.5. We considered two scenarios for the standard deviations δ_t that formed the diagonal matrix $\boldsymbol{\Delta}$. Scenario I specified $(\delta_1, \delta_2, \ldots, \delta_6)' = (1, \frac{5}{6}, \ldots, \frac{1}{6})'$ (heterogeneous variance situation), while scenario II specified equal variances with $\delta_t = 0.5$ ($t = 1, \ldots, 6$).

Once the complete dataset was generated, we excluded the treatment arms to create partially missing data as illustrated in Figure 5 under two mechanisms: (1) missing completely at random (MCAR) and (2) missing at random (MAR) with respect to absolute effects. We also considered two data structures for each missingness mechanism. For the first MCAR structure (denoted by MCAR1), we first kept all treatment 1 data (all 20 studies) and then kept each of the remaining treatments' data in a randomly chosen block of 4 studies, where the blocks did not overlap. Similarly, for the second MCAR structure (denoted by MCAR2), we also kept all treatment 1 data and then randomly kept data for treatments 2 to 6 data in blocks of 2, 2, 2, 2, and 12 studies, respectively, where again the blocks did not overlap. Under the MAR mechanism, the two data structures (denoted by MAR1 and MAR2) were specified in a similar manner. For both MAR1 and MAR2, we kept all treatment 1 data and ranked the studies in descending order by r_{k1}/n_{k1}. Then, for the MAR1 structure, we made treatment 3 available only in the first 4 studies (in this ordering), treatment 6 available in the next 4, and so on as in Figure 5. Similarly, for the MAR2 structure, we made treatment 3 available only in the first 2 studies, treatment 6 available in next 12, treatment 2 available in next 2, and so on as in Figure 5.

5.2 Simulation results

Table 2 summarizes the bias of the posterior mean ($\text{Bias}_{\bar{\mu}}$), the bias of the posterior median ($\text{Bias}_{\tilde{\mu}}$), the mean squared error (MSE) of the posterior median ($\text{MSE}_{\tilde{\mu}}$), and the coverage probability (CP) of the 95% CrI using the four priors under simulation scenario I with the four different missingness structures (MCAR1, MCAR2, MAR1, and MAR2). We evaluated the log odds ratio comparing treatments i and j (mLOR_{ij} and cLOR_{ij}), the AR of treatment t (p_t), the standard deviation for treatment t (δ_t), and the correlation between treatments i and j (ρ_{ij}).

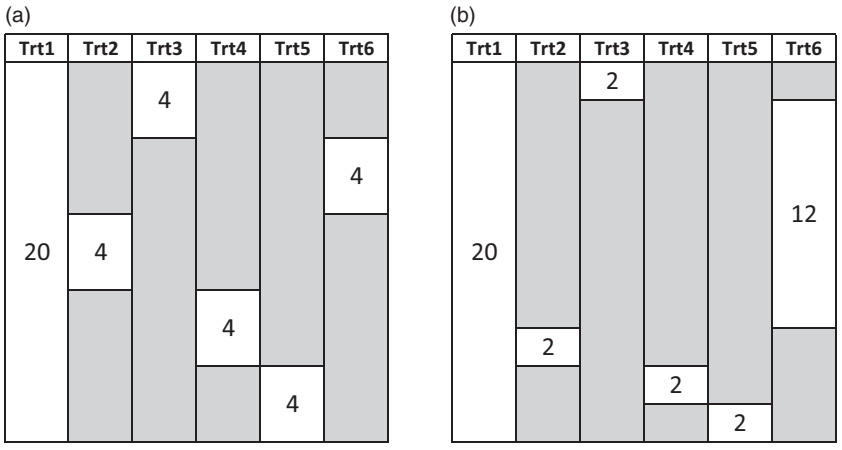

Figure 5. Missing data structures for simulation study: (a) MCAR1 and MAR1 and (b) MCAR2 and MAR2. The number in the white background indicates the observed clinical studies for each treatment, while the gray background indicates the corresponding treatment is not observed in these studies.
MAR: missing at random; MCAR: missing completely at random.

Due to space limits, instead of presenting the results for each treatment comparison, for each of Bias$\overline{\mu}$, Bias$_{\tilde{\mu}}$, and MSE$\tilde{\mu}$, we calculated the sum of the absolute value over all pairs of comparisons. For example, the entry in Table 2 with Bias$\overline{\mu}$ as the column and cLOR$_{ij}$ as the row was calculated as $\sum_{i\neq j}|\text{Bias}\overline{\mu}(\text{cLOR}_{ij})|$. To summarize the CPs, the corresponding value in Table 2 in column CP and row cLOR$_{ij}$ was calculated as $\sum_{i\neq j}(0.95-\text{CP}(\text{cLOR}_{ij}))_+$, where $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ if $x < 0$, i.e. the total shortfall in CP. Table 3 presents the simulation results under scenario II with similar summaries.

In both scenarios, using the UV and HHC priors, the posterior median was less biased than the posterior mean, especially for the MCAR2 and MAR2, in which the missingness structure was more unbalanced than MCAR1 and MAR1. However, using the IW and EV priors, the difference between these two point estimates was much smaller. With a weaker prior assumption, the UV and HHC priors may produce posterior distributions with larger skewness than the IW and EV priors when information was limited. Hence, for the remaining part, we focus on interpreting the posterior medians.

Comparing the HHC and UV priors, we could conclude that the HHC prior was much better in terms of bias and MSE for all parameters of interest under the four different missingness mechanisms. For the IW prior, estimates of the correlation and standard deviation were severely biased and had extremely poor CPs (though the MSE for standard deviations was the best among the four methods). Such biases had little influence on inference for log odds ratios and ARs when the data were MCAR, but for MAR1 and MAR2, the log odds ratio estimates produced by the IW prior were severely biased and much worse than those given by the HHC prior. The HHC and EV priors performed comparably in terms of bias and CP in scenario II, where the true variances were assumed equal, though the EV prior had better MSEs for all parameters than the HHC prior in scenario II. However, when the true variances were unequal (scenario I), the EV prior gave biased estimates and low CPs, while the HHC prior still gave estimates with reasonable biases and satisfactory CPs, especially under MAR.

Overall, the HHC prior provided the best estimates of log odds ratios and ARs among the four priors. The performance of the EV prior became worse when the homogeneity assumption was severely violated, and the IW prior had poor performance under MAR.

6 Summary and discussion

This article discussed different prior choices for the between-study standard deviations of multiple treatments in an NMA. We considered the traditional IW prior on the covariance matrix, a prior representing the UV assumption, and a prior representing the EV assumption, and we proposed the HHC prior. We compared these four priors using a real NMA. The results showed the superior performance of the HHC prior. Specifically, when the equal variance assumption was potentially violated, the HHC prior could still provide good results in terms of

Table 2. Simulation results for data generated under scenario I (heterogeneous variance) with four different missingness settings (MCAR1, MCAR2, MAR1, and MAR2), specifically bias of the posterior mean ($\text{Bias}_{\bar{\mu}}$), bias of the posterior median ($\text{Bias}_{\tilde{\mu}}$), mean squared error of the posterior median ($\text{MSE}_{\tilde{\mu}}$), and coverage probability (CP) of the 95% credible intervals for four different priors.

		IW				HHC				UV				EV			
Parameter	Truth	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP
MCAR1																	
cLOR_{ij}	.	0.52	0.50	2.58	0.00	0.39	0.32	2.48	0.01	0.52	0.44	2.79	0.00	0.74	0.75	2.69	0.00
mLOR_{ij}	.	0.81	0.80	2.22	0.00	0.40	0.07	2.31	0.02	1.67	1.11	2.59	0.00	1.70	1.69	2.34	0.00
p_t	.	0.06	0.02	0.00	0.02	0.06	0.02	0.00	0.03	0.14	0.07	0.01	0.00	0.09	0.06	0.01	0.05
δ_t	.	1.02	0.90	0.34	1.22	0.62	0.15	0.49	0.02	2.25	1.29	1.21	0.12	2.03	1.97	1.09	3.31
mLOR_{35}	−1.40	−0.03	−0.04	0.15	0.99	0.05	0.01	0.16	0.97	0.21	0.15	0.20	0.99	−0.04	−0.04	0.14	0.99
p_5	0.06	0.01	0.00	0.00	0.96	0.01	0.00	0.00	0.95	0.03	0.01	0.00	0.96	0.01	0.01	0.00	0.97
ρ_{35}	0.50	−0.46	−0.45	0.22	1.00	−0.04	−0.02	0.04	0.98	−0.01	0.02	0.04	0.98	−0.13	−0.11	0.05	0.97
MCAR2																	
cLOR_{ij}	.	0.74	0.68	5.34	0.02	0.96	0.64	5.44	0.00	0.85	0.78	6.63	0.00	0.87	0.87	4.76	0.01
mLOR_{ij}	.	0.95	1.02	4.46	0.00	0.97	0.53	4.43	0.00	4.17	2.82	4.58	0.00	1.91	1.90	4.10	0.10
p_t	.	0.09	0.04	0.01	0.02	0.13	0.04	0.01	0.01	0.34	0.15	0.02	0.00	0.11	0.07	0.01	0.07
δ_t	.	1.10	0.93	0.31	1.44	1.69	0.19	1.06	0.00	4.74	3.22	3.71	0.11	1.86	1.82	0.95	3.23
mLOR_{35}	−1.40	−0.09	−0.10	0.35	0.99	0.02	−0.05	0.34	0.99	0.26	0.20	0.35	1.00	−0.10	−0.10	0.32	0.99
p_5	0.06	0.01	0.00	0.00	0.98	0.03	0.01	0.00	0.96	0.08	0.03	0.00	0.98	0.01	0.01	0.00	0.98
ρ_{35}	0.50	−0.49	−0.49	0.24	1.00	−0.05	−0.03	0.04	1.00	−0.05	−0.03	0.04	1.00	−0.21	−0.21	0.07	0.94
MAR1																	
cLOR_{ij}	.	2.77	2.78	3.61	0.00	1.80	0.91	3.42	0.00	8.88	7.02	10.86	0.03	6.44	6.99	8.10	0.73
mLOR_{ij}	.	2.93	2.70	2.90	0.00	0.91	0.60	2.44	0.00	4.04	3.75	4.48	0.05	5.05	5.55	5.40	0.70
p_t	.	0.08	0.06	0.01	0.01	0.10	0.03	0.01	0.01	0.28	0.17	0.02	0.07	0.20	0.19	0.03	0.19
δ_t	.	1.27	1.02	0.41	1.25	1.01	0.30	0.61	0.01	3.89	2.72	3.09	0.33	2.27	2.18	1.33	3.35
mLOR_{35}	−1.40	0.44	0.42	0.39	0.99	−0.01	0.02	0.22	0.99	−0.47	−0.44	0.48	0.99	−0.80	−0.90	1.14	0.85
p_5	0.06	0.04	0.02	0.00	0.98	0.01	0.01	0.00	0.97	0.01	−0.00	0.00	0.99	−0.00	−0.01	0.00	0.96
ρ_{35}	0.50	−0.49	−0.49	0.25	1.00	−0.03	0.01	0.04	1.00	0.20	0.29	0.11	0.94	0.12	0.18	0.08	0.91
MAR2																	
cLOR_{ij}	.	4.90	4.81	7.17	0.00	4.29	0.80	6.41	0.00	14.93	10.52	21.84	0.00	6.81	7.50	11.34	0.38
mLOR_{ij}	.	4.60	4.57	5.92	0.00	1.69	0.99	4.93	0.01	7.70	6.75	9.04	0.00	5.86	6.46	8.40	0.48
p_t	.	0.09	0.09	0.01	0.01	0.19	0.04	0.01	0.01	0.54	0.33	0.06	0.01	0.25	0.24	0.04	0.11
δ_t	.	1.26	1.00	0.33	1.58	1.95	0.19	1.07	0.01	5.26	3.90	4.90	0.15	2.10	2.02	1.16	3.24
mLOR_{35}	−1.40	0.62	0.60	0.65	1.00	−0.12	0.01	0.35	1.00	−0.88	−0.84	1.15	1.00	−0.91	−1.02	1.61	0.90
p_5	0.06	0.07	0.03	0.00	0.99	0.02	0.01	0.00	0.99	0.03	0.00	0.00	1.00	0.00	−0.01	0.00	0.95
ρ_{35}	0.50	−0.50	−0.50	0.25	1.00	−0.03	−0.00	0.03	1.00	0.10	0.18	0.06	1.00	0.08	0.14	0.06	0.95

Specifically as an example, the table entry in column $\text{Bias}_{\bar{\mu}}$ and row cLOR_{ij} was defined as $\sum_{i \neq j} |\text{Bias}_{\bar{\mu}}(\text{cLOR}_{ij})|$. Similarly, the table entry in column CP and row cLOR_{ij} was defined as $\sum (0.95 - \text{CP}(\text{cLOR}_{ij}))_+$.

IW: inverse-Wishart; HHC: hierarchical half-Cauchy; CP: coverage probability; MCAR: missing completely at random; MSE: mean squared error; MAR: missing at random.

Table 3. Simulation results comparing data generated under scenario II (homogeneous variance) with four different missingness settings (MCAR1, MCAR2, MAR1, and MAR2).

		IW				HHC				UV				EV			
Parameter	Truth	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\bar{\mu}}$	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP
MCAR1																	
cLOR_{ij}	.	0.41	0.37	2.04	0.00	0.31	0.22	1.95	0.00	0.50	0.39	2.22	0.00	0.24	0.23	1.80	0.01
mLOR_{ij}	.	0.19	0.11	1.74	0.00	0.29	0.08	1.80	0.00	1.50	0.88	2.07	0.00	0.18	0.18	1.64	0.01
p_t	.	0.05	0.02	0.00	0.00	0.05	0.02	0.00	0.02	0.13	0.06	0.00	0.00	0.02	0.01	0.00	0.02
δ_t	.	0.74	0.51	0.09	0.00	0.39	0.10	0.36	0.00	2.11	1.12	0.97	0.11	0.15	0.10	0.05	0.00
mLOR_{35}	−1.44	0.00	−0.00	0.14	0.99	0.02	−0.00	0.14	0.97	0.14	0.09	0.16	0.99	−0.02	−0.02	0.13	0.95
p_5	0.05	0.01	0.00	0.00	0.98	0.01	0.00	0.00	0.96	0.03	0.01	0.00	0.97	0.00	0.00	0.00	0.96
ρ_{35}	0.50	−0.48	−0.47	0.23	1.00	−0.02	0.00	0.04	0.99	0.00	0.03	0.05	0.99	−0.06	−0.04	0.05	0.97
MCAR2																	
cLOR_{ij}	.	0.65	0.60	3.89	0.00	0.70	0.53	3.76	0.00	0.80	0.70	5.15	0.00	0.52	0.50	3.39	0.19
mLOR_{ij}	.	0.36	0.28	3.30	0.00	1.00	0.35	3.29	0.00	4.48	3.02	4.09	0.00	0.45	0.44	3.08	0.17
p_t	.	0.08	0.03	0.01	0.00	0.12	0.03	0.01	0.02	0.33	0.15	0.02	0.00	0.04	0.02	0.01	0.05
δ_t	.	0.94	0.62	0.11	0.00	1.45	0.07	0.58	0.03	4.82	3.25	3.65	0.11	0.14	0.08	0.05	0.04
mLOR_{35}	−1.44	−0.02	−0.03	0.35	0.99	0.03	−0.03	0.33	1.00	0.28	0.22	0.36	1.00	−0.07	−0.07	0.32	0.93
p_5	0.05	0.01	0.00	0.00	0.98	0.03	0.00	0.00	0.96	0.08	0.03	0.00	0.98	0.00	0.00	0.00	0.93
ρ_{35}	0.50	−0.50	−0.50	0.25	1.00	−0.02	−0.00	0.05	0.99	−0.02	0.01	0.05	0.99	−0.05	−0.03	0.05	0.98
MAR1																	
cLOR_{ij}	.	3.25	3.24	2.89	0.00	0.86	0.17	2.40	0.00	7.51	5.56	7.32	0.00	0.53	0.40	1.97	0.02
mLOR_{ij}	.	3.40	3.30	2.65	0.00	0.43	0.43	2.07	0.00	3.72	3.19	3.70	0.01	0.56	0.42	1.78	0.01
p_t	.	0.08	0.07	0.00	0.00	0.07	0.02	0.01	0.02	0.25	0.15	0.02	0.03	0.02	0.01	0.00	0.02
δ_t	.	0.73	0.48	0.08	0.00	0.51	0.07	0.40	0.00	3.25	2.07	2.02	0.29	0.14	0.08	0.04	0.00
mLOR_{35}	−1.44	0.53	0.53	0.41	0.97	0.01	0.05	0.20	0.99	−0.51	−0.48	0.53	0.99	0.09	0.07	0.16	0.98
p_5	0.05	0.03	0.02	0.00	0.97	0.01	0.01	0.00	0.97	0.01	−0.00	0.00	1.00	0.01	0.00	0.00	0.97
ρ_{35}	0.50	−0.50	−0.50	0.25	1.00	−0.05	−0.02	0.04	1.00	0.18	0.28	0.10	0.98	−0.10	−0.08	0.05	1.00
MAR2																	
cLOR_{ij}	.	4.64	4.62	5.04	0.00	2.95	0.37	4.38	0.00	12.74	8.38	15.75	0.00	0.73	0.58	3.22	0.00
mLOR_{ij}	.	4.41	4.44	4.45	0.00	1.11	0.71	3.61	0.00	6.67	5.30	7.03	0.00	0.76	0.59	2.92	0.01
p_t	.	0.08	0.09	0.01	0.00	0.15	0.02	0.01	0.02	0.49	0.28	0.05	0.00	0.03	0.01	0.01	0.01
δ_t	.	0.93	0.60	0.10	0.00	1.54	0.07	0.57	0.02	5.24	3.77	4.49	0.13	0.14	0.08	0.05	0.00
mLOR_{35}	−1.44	0.71	0.71	0.74	0.99	−0.01	0.10	0.39	1.00	−0.67	−0.62	0.91	1.00	0.10	0.07	0.30	0.96
p_5	0.05	0.04	0.03	0.00	0.99	0.02	0.01	0.00	0.99	0.03	0.01	0.00	1.00	0.01	0.00	0.00	0.96
ρ_{35}	0.50	−0.50	−0.50	0.25	1.00	−0.05	−0.02	0.04	1.00	0.06	0.12	0.06	0.99	−0.08	−0.06	0.05	0.99

The bias of posterior mean ($\text{Bias}_{\bar{\mu}}$), the bias of posterior median ($\text{Bias}_{\tilde{\mu}}$), the mean squared error of posterior median ($\text{MSE}_{\tilde{\mu}}$), and the coverage probability (CP) of the 95% credible intervals were summarized for four different priors. Specifically as an example, the value in the table with $\text{Bias}_{\bar{\mu}}$ as column and cLOR_{ij} as row was defined as $\sum_{i\neq j}|\text{Bias}_{\bar{\mu}}(\text{cLOR}_{ij})|$. Moreover, the value in the table with column CP and row cLOR_{ij} was defined as $\sum_{i\neq j}(0.95 - \text{CP}(\text{cLOR}_{ij}))_+$.

IW: inverse-Wishart; HHC: hierarchical half-Cauchy; CP: coverage probability; MSE: mean squared error; MCAR: missing completely at random; MAR: missing at random.

deviance and DIC, while the UV prior overestimated the variances of treatments that had limited information. On the other hand, the EV prior may provide biased estimates for variances and did not fit the data well. In addition to the analyses presented here, we did a sensitivity analysis to explore the prior's impact on μ_t. Specifically, as suggested by Gelman et al.[36] and Ghosh et al.,[37] we considered the Student's t-prior $t_7(10)$ on fixed effects μ_t, which has 7 degrees of freedom and location parameter 10. The results were similar to those in Section 4; the DICs were almost unchanged at 92.55, 90.75, and 94.43 for the UV, HHC, and EV priors, respectively.

We also compared the performance of the different priors using simulation studies with various settings and missing treatment structures. Table 4 summarizes the pros and cons of these methods. The UV prior leads to biased estimates, and the CrIs did not have nominal CP when B_t was small (≤ 4). The IW prior could not estimate correlations and standard deviations accurately, which resulted in biased log odds ratios and ARs under the MAR mechanism. The EV prior could produce unbiased estimates when the true variances were equal, but when the homogeneous variance assumption was severely violated, it gave biased estimates and 95% CrIs with poor

Table 4. Pros and cons of the four different models.

		Performance					
		MCAR			MAR		
			Small B_t exists			Small B_t exists	
Model	Computing burden	All B_ts are large	δ_ts are similar	δ_ts differ	All B_ts are large	δ_ts are similar	δ_ts differ
IW	Small	Good	Good	Good	Bad	Bad	Bad
UV	Large	Good	Bad	Bad	Good	Bad	Bad
EV	Large	Depends	Good	Bad	Depends	Good	Bad
HHC	Large	Good	Good	Good	Good	Good	Good

IW: inverse-Wishart; HHC: hierarchical half-Cauchy; MCAR: missing completely at random; MAR: missing at random.

coverage. The HHC prior generally had the best performance among the four priors in terms of estimating relative effects and absolute effects. It produced almost unbiased results and satisfactory CP using a weaker assumption than the EV prior.

This article focused on shrinking standard deviations in the AB-NMA with binary outcomes; many extensions are possible. First, shrinkage is feasible for AB-NMA with other types of outcomes. For instance, the observed data for NMA with continuous outcomes are $D_k = \{(\overline{y}_{kt}, s_{kt}, n_{kt}), t \in A_k\}$, where $\overline{y}_{kt}$, s_{kt}, and n_{kt} are the sample mean, its standard error, and the sample size for the t^{th} treatment in the k^{th} study, respectively. The model for AB-NMA with continuous outcomes is Zhang et al.[27]

$$\begin{array}{l} \text{Level I}: \overline{y}_{kt} \sim N(\theta_{kt}, s_{kt}^2/n_{kt}),\ t \in A_k,\ k = 1, \ldots, K; \\ \text{Level II}: (\theta_{k1}, \ldots, \theta_{kT})' \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{array} \tag{7}$$

where θ_{kt} is the underlying mean outcome of the t^{th} treatment in the k^{th} study. Comparing equations (1) and (7), the difference is on Level I (within study); because shrinkage is applied to δ_t, $t = 1, \ldots, K$, at the between-study level, it can be applied to AB-NMA with various outcomes simply by modifying the likelihood and link functions. Nevertheless, additional case studies and simulations need to be performed to examine the performance of the variance shrinkage method in other settings.

Second, the shrinkage method could potentially be applied to the CB-NMA. For CB-NMA, we need to focus on the standard deviations of contrasts, instead of standard deviations of treatment effects as in AB-NMA. Also, the triangle inequalities on contrast standard deviations[9] could complicate prior specifications.

Third, the HHC is just one choice of prior for inducing shrinkage. Other priors, such as the hierarchical inverse-gamma (HIG) prior, denoted by $\text{HIG}(\epsilon_l, \epsilon_u)$, could be considered. The HIG prior's density is $p(\delta_t^2|\beta) \propto \text{IG}(\alpha, \beta)$ with α fixed at 1 and β following the uniform distribution $U(\epsilon_l, \epsilon_u)$. Like the HHC prior, the HIG prior (e.g. with $\epsilon_l = 0$ and $\epsilon_u = 1$) can adaptively achieve a balance between an informative prior with high density near zero, such as $\text{IG}(1, 0.1)$, and a prior that may overestimate a variance with true value close to zero, such as $\text{IG}(1, 1)$ (see Figure 4). We compared $\text{HIG}(0, 1)$ with other four methods (see Online Appendix B) and found that the performance of HIG was better than IW, EV, and UV methods but slightly worse than the HHC method in the simulation studies. In addition, based on WAIC and DIC, HIG and HHC methods were similar in analyzing the case study.

Fourth, while we chose $\epsilon_l = 0$ and $\epsilon_u = 5$ in the uniform prior for the HHC's hyperparameter a, this choice needs careful justification in practice. For example, the lower bound of the uniform prior ϵ_l places an upper bound on the informativeness of the HHC prior. Specifically, $\epsilon_l = 0.1$ might be a better choice than $\epsilon_l = 0$ since HC(0.1) is less informative than HC(0.001).

Finally, the variance shrinkage method may still involve some hidden assumptions about variances. It may be critical to assess the implications of assuming that different treatment variances share a common distribution with somewhat arbitrarily chosen hyperparameters, as in the HHC prior. On the other hand, AB-NMA can naturally include single-arm studies when they are available to make inference with more information. However, including single-arm studies in an NMA may require additional assumptions about the mean and variance parameters. Therefore, it may be more sensible to allow the between-study variances to be similar (i.e. sharing a common

distribution) but not exactly the same in multiarm (≥ 2) studies versus single-arm studies. Therefore, methods for combining single-arm and multiple-arm studies should be further examined.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by NIH NLM R01LM012982.

ORCID iD

Zhenxun Wang https://orcid.org/0000-0002-1636-7240

Supplemental material

Supplemental material, including code to support our findings for this article, has been submitted to the journal and will be available online.

References

1. Trinder L and Reynolds S. (eds) A critical appraisal of evidence-based practice. In: *Evidence-based practice*. Hoboken, NJ: Blackwell Science Ltd, 2000, pp.212–241.
2. Egger M, Davey Smith G and Altman DG (eds) *Systematic reviews in health care*. London: BMJ Publishing Group, 2001.
3. Mills EJ, Ioannidis JPA, Thorlund K, et al. How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA* 2012; **308**: 1246–1253.
4. Lu G and Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; **23**: 3105–3124.
5. Zhang J, Carlin BP, Neaton JD, et al. Network meta-analysis of randomized clinical trials: reporting the proper summaries. *Clin Trials* 2014; **11**: 246–262.
6. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; **50**: 683–691.
7. Rücker G and Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015; **15**: 58.
8. Lu G and Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006; **101**: 447–459.
9. Lu G and Ades AE. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009; **10**: 792–805.
10. Hong H, Chu H, Zhang J, et al. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods* 2016; **7**: 6–22.
11. Zhang J, Chu H, Hong H, et al. Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness. *Stat Methods Med Res* 2017; **26**: 2227–2243.
12. Dias S and Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. *Res Synth Methods* 2016; **7**: 23–28.
13. Hong H, Chu H, Zhang J, et al. Rejoinder to the discussion of "A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons", by S. Dias and A.E. Ades. *Res Synth Methods* 2016; **7**: 29–33.
14. Béliveau A, Goring S, Platt RW, et al. Network meta-analysis of disconnected networks: how dangerous are random baseline treatment effects? *Res Synth Methods* 2017; **8**: 465–474.
15. Lin L, Chu H and Hodges JS. Sensitivity to excluding treatments in network meta-analysis. *Epidemiology* 2016; **27**: 562–569.
16. Nikolakopoulou A, Chaimani A, Veroniki AA, et al. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS One* 2014; **9**: e86754.
17. Wang Z, Lin L, Hodges JS, et al. The impact of covariance priors on arm-based Bayesian network meta-analyses with binary outcomes. *Stat Med*. Epub ahead of print 3 June 2020. DOI:10.1002/sim.8580.
18. Dias S, Sutton AJ, Ades AE, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making* 2013; **33**: 607–617.
19. James W and Stein C. Estimation with quadratic loss. In: Kotz S and Johnson NL (eds) *Breakthroughs in Statistics. Springer Series in Statistics (Perspectives in Statistics) Springer series in statistics*. New York: Springer, 1992, pp.443–460.

20. Hwang JTG, Qiu J and Zhao Z. Empirical Bayes confidence intervals shrinking both means and variances. *J R Stat Soc Ser B Stat Methodol* 2009; **71**: 265–285.
21. Hwang JG and Liu P. Optimal tests shrinking both means and variances applicable to microarray data analysis. *Stat Appl Genet Mol Biol* 2010; **9**: 36.
22. Zhao Z. Double shrinkage empirical Bayesian estimation for unknown and unequal variances. *Stat Interface* 2010; **3**: 533–541.
23. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *Lancet* 2012; **380**: 2095–2128.
24. Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *Lancet* 2012; **380**: 2197–2223.
25. Martin P. Stroke units: an evidence based approach. *J Neurol Neurosurg Psychiatry* 1999; **66**: 412–412.
26. Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. *Cochrane Database Syst Rev* 2007; **4**: Art. No.: CD000197.
27. Zhang J, Fu H and Carlin BP. Detecting outlying trials in network meta-analysis. *Stat Med* 2015; **34**: 2695–2707.
28. Barnard J, McCulloch R and Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Stat Sin* 2000; **10**: 1281–1311.
29. Lin L, Zhang J, Hodges JS, et al. Performing arm-based network meta-analysis in R with the pcnetmeta package. *J Stat Softw* 2017; **80**: 1–25.
30. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 2006; **1**: 515–534.
31. Zeger SL, Liang KY and Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; **44**: 1049–1060.
32. Salanti G, Ades A and Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; **64**: 163–171.
33. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol* 2002; **64**: 583–639.
34. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res* 2010; **11**: 3571–3594.
35. Lunn D, Jackson C, Best N, et al. *The BUGS book*. New York: Chapman and Hall/CRC.
36. Gelman A, Jakulin A, Pittau MG, et al. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2008; **2**: 1360–1383.
37. Ghosh J, Li Y and Mitra R. On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Anal* 2018; **13**: 359–383.