STATISTICAL METHODS FOR ARM-BASED BAYESIAN NETWORK META-ANALYSIS

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

ZHENXUN WANG

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISED BY DR. HAITAO CHU

June, 2020

Acknowledgements

I would like to express my sincere gratitude to my advisor, Prof. Haitao Chu for his guidance and support in my PhD studies. Prof. Haitao Chu is the best advisor and friend a student could have. Beyond statistical, epidemiological, and clinical knowledge, he showed amazing intuition about research problems. I am glad that Haitao taught me a lot about fundamentals of scientific research, from proposing meaningful research questions to communicating information to other scientists efficiently.

I owe a debt of gratitude to Prof. Lifeng Lin for his valuable time, insightful remarks and careful attention to detail on my dissertation. I am grateful to my dissertation committee, Professors James Hodges, Thomas Murray, and Richard MacLehose. They provided numerous helpful suggestions on research questions. Prof. James Hodges also provided great guidance on both scientific writing and statistical modeling. Moreover, I would like to thank Professors Kyle Rudser, Cavan Reilly, Wei Pan and Xianghua Luo for their support and guidance during my graduate assistantships. I treasure the rewarding experience working in the PLUS-SDCC, where I learned how to collaborate and communicate with non-statisticians. In addition, the University of Minnesota's Division of Biostatistics, Prof. Haitao Chu, and the Health Policy Statistics Section of the American Statistical Association provided funds for me to present work at different conferences.

Finally, I would like to thank my family, especially my dad Zhaoliang Wang and my mom Meijun Wu, for their encouragement, trust and support throughout my life.

Dedication

To my parents, Zhaoliang Wang and Meijun Wu, for their endless support, trust, and love.

Abstract

Network meta-analysis (NMA) is a recently developed tool to combine and contrast direct and indirect evidence in systematic reviews of multiple treatments. Compared to traditional pairwise meta-analysis, it can improve statistical efficiency and reduce certain biases. Unlike the contrast-based NMA approach, which focuses on estimating relative effects such as odds ratios, the arm-based (AB) NMA approach can estimate absolute effects (such as overall treatment-specific event rates), which are arguably more useful in medicine and public health, as well as relative effects. Bayesian analyses with the arm-based (AB) network meta-analysis (NMA) model require researchers to specify a prior distribution for the covariance matrix of the treatment-specific event rates on a transformed scale, e.g., the treatment-specific log-odds when a logit transformation is used. Specifically, in AB-NMA, standard deviations of study-specific log-odds are needed to derive treatment-specific overall effects, while accurate estimation of correlation coefficients is critical for borrowing information across treatments. However, partially due to a lack of information, estimation of correlation coefficients and variances can be biased and unstable if we use a conjugate prior (e.g., inverse-Wishart (IW) distribution) for the covariance matrix.

To address the first challenge of accurately estimating correlation coefficients, several separation strategies (i.e., separate priors on variances and correlations) can be considered. To study the IW prior's impact on AB-NMA and compare it with separation strategies, we did simulation studies under different missing-treatment mechanisms. A separation strategy with appropriate priors for the correlation matrix (e.g., equal correlations) performs better than the IW prior. It is thus recommended as the default vague prior in the AB approach. We also re-analyzed three case studies and illustrated the importance, when performing AB-NMA, of sensitivity analyses with different prior specifications on variances.

To address the second challenge of variance estimation, we propose two approaches. We first introduce a variance shrinkage method. Specifically, we assume different treatment-specific variances share a common prior with unknown hyper-parameters. 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 re-analyzing an NMA of organized inpatient care interventions for stroke. Comprehensive simulations investigate the impact of different variance assumptions on statistical inference, and these simulation results show that the variance shrinkage method provides better estimates of log odds ratios and absolute risks.

In the second approach to improving variance estimation, we consider borrowing information from single arm studies (variance extrapolation) in AB-NMA. AB-NMA model can naturally incorporate information from single-arm studies about means and variances. However, single-arm studies and randomized clinical trials (RCTs) may have different study populations and study quality, so that assuming they are exchangeable may be inappropriate. We present a novel *commensurate prior on variance* (CPV) method to borrow variance (rather than mean) information from single-arm studies in an arm-based (AB) Bayesian NMA. We illustrate the advantages of this CPV method by reanalyzing an NMA of immune checkpoint inhibitors in cancer patients. Comprehensive simulations investigate the impact on statistical inference of including single-arm studies. The simulation results show that the CPV method provides efficient and robust estimation even when the two sources of information are moderately inconsistent.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Evidence-based practice (EBP) is a powerful theoretical framework to connect study findings to a profession's body of knowledge [2]. 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 estimates of effects, can address certain biases (e.g., reporting bias and small-study effects) of estimated treatment effects [3].

Traditionally, a meta-analysis of randomized controlled trials compares only two treatments, typically an intervention and a control. With new treatments emerging, network meta-analysis (NMA) was developed to simultaneously compare multiple (more than two) interventions. Compared to pairwise comparison from a traditional meta-analysis, NMA can gain precision by considering both direct and indirect comparisons and has the potential to rank regimens more explicitly [4]. For instance, if three treatments A, B, and C are available for a certain disease, then in NMA, comparing treatments A and C provides direct evidence about A versus C while comparing A versus B plus B versus C offers indirect evidence. Both Bayesian hierarchical approaches [5–7] and frequentist methods [8, 9] have been proposed for NMA; this thesis focuses on Bayesian methods because they have been widely applied [8, 9].

Two widely-used Bayesian hierarchical approaches have been considered, the contrast-based (CB) approach [5, 6, 10] and the arm-based (AB) approach [7, 11, 12]. 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 [13–15]. The primary advantage of AB-NMA is that it naturally estimates absolute risks and absolute risk differences. Absolute risks 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 [16]. In the remaining chapter, I will discuss some potential problems in Bayesian analyses of the AB-NMA model and some solutions to them.

1.1 Covariance priors

Both the AB and CN approaches involve estimating the covariance matrix of random effects and choosing a prior distribution for it, which are generally difficult in Bayesian analysis: the number of parameters in a covariance matrix increases rapidly with the dimension of the matrix, and these parameters are constrained because the matrix must be non-negative definite [17]. Also, the parameters of the two models are distinct. In the CB approach, if the total number of treatments in the NMA is T, then $T-1$ variance parameters need to be estimated for contrasts and $(T-1)(T-2)/2$ parameters for correlations between contrasts but the $T-1$ variances are constrained by triangle inequalities [6], which complicate prior specification for heterogeneous variances. The AB approach estimates more parameters (i.e., a T-dimensional covariance matrix for variances of absolute effects and correlations between them), but does so without constraints other than positive definiteness. Finally, each study generally includes only a small portion of the NMA's treatments, generally based on results of previous trials. This selection of treatments produces missing data (treatments excluded from a trial), which affects the AB and CB approaches differently because they use different exchangeability assumptions: CB models require contrasts to be missing at random (MAR) [18] while AB models require treatments to be MAR [11]. Such differences could have distinct effects on estimates of the covariance matrix.

To model the variance structure of contrasts in the CB approach, Lu and Ades [6] used an ancillary representation to circumvent the triangle inequalities and compared different prior specifications using case studies. However, in the AB approach little attention has been paid to the choice of priors for covariance matrices and their influence on the results. The obvious choice is the conjugate inverse-Wishart prior, which was used by Zhang et al. [7] and Hong et al. [11] However, this prior has some problems: the marginal distribution of the variances has low density in the region near zero [19] and the prior imposes a dependency between the variances and the correlations [20], which may cause the correlations to be underestimated in certain situations. Correlations are critical for borrowing strength across treatment arms to estimate treatment effects efficiently and to reduce potential bias. Hence, in Chapter 2, we consider and compare different covariance priors in AB-NMA using case studies and simulation studies.

1.2 Lack of information

Generally, randomized controlled trials (RCTs) with blinded outcome assessment offer high-quality, reliable evidence for statistical analyses [3] and are preferred for inclusion in meta-analyses. Partly because of strict screening processes, however, nearly half of the meta-analyses in the Cochrane Database of Systematic Reviews contain only two or three studies [21]. It is challenging to select an appropriate method for meta-analysis with only a few studies (≤ 5) balancing statistical power and nominal coverage probability [22]. Similarly, in a survey of 186 NMAs, nearly 40% of treatments were included in four or fewer trials, and the median number of trials per comparison was 2 (interquartile range, 1–4) [1]. 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 [1]. This phenomenon may produce difficulties in accurately estimating correlations among treatments in the AB approach, and what we propose in Chapter 2 could be a remedy. In the CB approach, lack of information may cause variances of certain contrasts to be overestimated [6]. 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. [1] Descriptive statistics

of these 42 networks (Figure 1.1) show that nearly 40% of treatments in these NMAs are included in 4 or fewer clinical studies. Because of this, variances of outcomes for individual treatments (absolute effects) in AB-NMA are difficult to be estimated.

1.2.1 Variance shrinkage

To overcome this problem, both the AB and CB approaches often make additional assumptions. For example, in the CB model, Dias et al. [23] 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 [24] and the double shrinkage estimator [25–27], Chapter 3 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.

1.2.2 Variance extrapolation

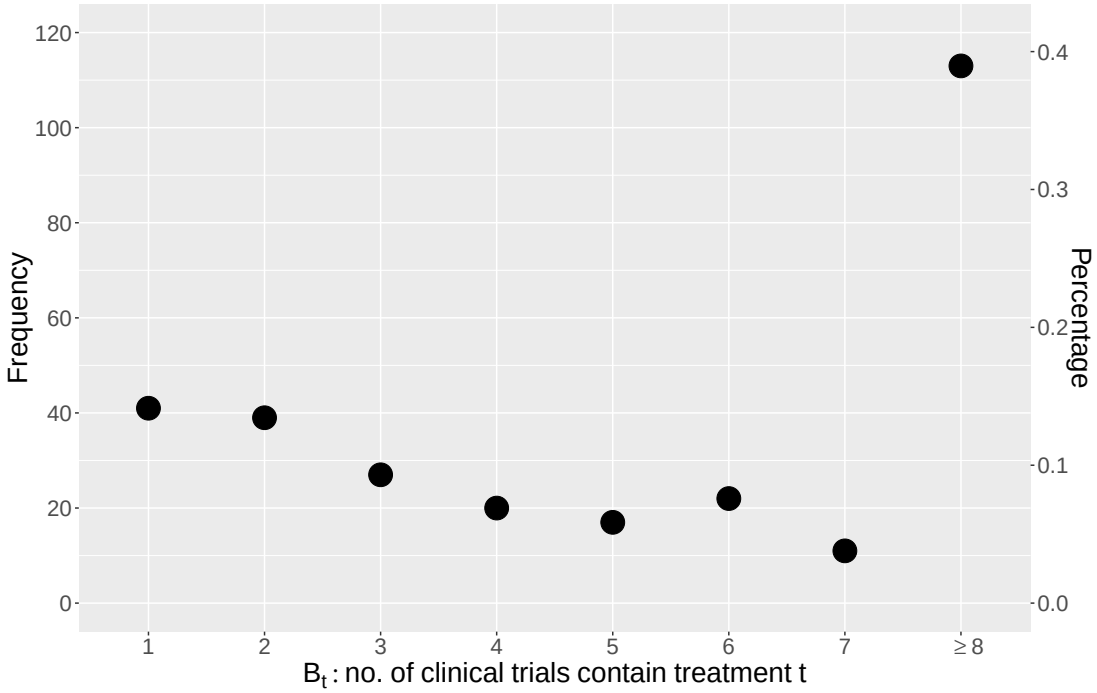
Another strategy could use extrapolation. When information is sparse in a targeted population, information borrowing is a useful technique for incorporating an external data source to improve statistical estimation. It has been widely used in RCTs, incorporating historical controls when diseases are rare or patient populations are small [28–31]. While the history of borrowing external information in evidence synthesis dates back to the 1990s, when Begg and Pilote [32] and Li and Begg [33] tried to combine results from controlled and uncontrolled studies using a frequentist approach, only recently have we witnessed a surge of publications on this topic. For example, Zhang et al. [34] proposed methods for a meta-analyses to adaptively combine RCTs and single-arm studies, while Röver et al. [35] used Bayesian model averaging to borrow adult evidence in pediatric meta-analysis, which generally has fewer studies available. Efthimiou et al. [36] recently proposed approaches to combining randomized and non-randomized evidence in

CB-NMA. Turner et al. [37] introduced four different informative priors for multiple heterogeneity variances in CB-NMA, and Leahy et al. [38] tried to incorporate single-arm evidence in a CB-NMA using aggregate-level covariate matching. There are also AB-NMA methods to synthesize aggregate and individual patient data [39].

So far, however, very little attention has been paid to including single-arm studies in an AB-NMA, partly because only AB-NMAs can include them and there is an ongoing debate about the relative merits of CB-NMA and AB-NMA [13, 14, 40]. The AB approach has the potential to estimate absolute risks, which are necessary to calculate cost-effectiveness when making decisions about drug coverage. As mentioned above, however, scant information is so prevalent in NMA that it is difficult to estimate the standard deviations across studies of treatment-specific effects (e.g., the log odds, if the logit transformation is used in AB-NMA with binary outcomes). Although a homogeneous variance assumption or variance shrinkage methods can help, these methods require some strong assumptions about variances. Hence, in Chapter 4, we develop methods that can incorporate extra evidence from single-arm studies in AB-NMA, to provide better estimation.

Finally, Chapter 5 summarizes the major findings in this thesis and introduced directions for future research.

Figure 1.1: Dot plot describing 42 NMAs with binary outcomes, from a total of 186 NMAs investigated by Nikolakopoulou et al. [1] 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 4 or fewer clinical trials.



Chapter 2

The Impact of Covariance Priors on Arm-Based Bayesian Network Meta-Analyses with Binary Outcomes

2.1 Introduction

As mentioned in Section 1.1, the commonly-used conjugate prior for the covariance matrix, the inverse-Wishart (IW) distribution, has several limitations in AB-NMA. Alternatives have been proposed, such as the scaled inverse Wishart [41], a separation strategy [17], the Cholesky decomposition [42], and the LKJ prior for a correlation matrix [43]. This chapter compares, for binary outcomes, the influence of selected priors for the AB model on the estimation of the log odds ratios of treatment comparisons and the correlations between treatment-specific log-odds, under different missing-treatment mechanisms. Based on these comparisons, we aim to recommend appropriate vague priors for the AB model's covariance matrix.

The rest of this chapter is organized as follows. Section 2.2 describes AB approaches for NMA, followed by Section 2.3 on Bayesian analysis, focusing on priors for the covariance matrix. Section 2.4 presents simulation studies and results, followed by three case

studies in Section 2.5. Section 2.6 presents our main conclusions with a brief discussion.

2.2 The arm-based network meta-analysis

2.2.1 Notation

Suppose we have collected K studies comparing a total of T treatments and each study contains only a subset of the T treatments. Let A_k $(k = 1, \ldots, K)$ be the subset of treatments investigated in the k^{th} study. If the number of elements in the set A_k (denoted by $|A_k|$) is larger than 2, then study k is called multi-armed. Most randomized clinical trials are two-arm studies with $|A_k| = 2$. Let $D_k = \{(r_{kt}, n_{kt}), t \in A_k\}$ be the data collected in the k^{th} study, where r_{kt} and n_{kt} are the numbers of events and total subjects in the t^{th} treatment group in the k^{th} study. Finally, let p_{kt} be the true probability of an event (i.e., absolute risk) for the t^{th} treatment in the k^{th} study.

2.2.2 The arm-based approach

Zhang et al. [7] proposed the following arm-based NMA model:

$$\begin{aligned} & r_{kt} \sim \text{Binomial}(n_{kt}, p_{kt}), \quad t \in A_k, k = 1, \ldots, K; \\ & \text{logit}(p_{kt}) = \mu_t + \nu_{kt}; \\ & (\nu_{k1}, \ldots, \nu_{kT})' \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \tag{2.1}$$

where μ_t represents the overall fixed effect of treatment t and the vector $(\nu_{k1}, \ldots, \nu_{kT})'$ is a random effect specific to study k, following the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ having dimension T. Here, $\boldsymbol{a}'$ denotes the transpose of the vector $\boldsymbol{a}$. We used logit instead of probit transformation to estimate both marginal and conditional log odds between treatments. Let δ_t be the between-study standard deviation of log-odds for treatment t, i.e., the square root of the t^{th} diagonal element in the covariance matrix $\boldsymbol{\Sigma}$. Here, $\boldsymbol{\Sigma}$ can also be written as $\boldsymbol{\Delta}\mathbf{P}\boldsymbol{\Delta}$ with $\boldsymbol{\Delta}$ a diagonal matrix having standard deviation δ_t as its t^{th} diagonal element, and correlation matrix $\mathbf{P}$ with entries ρ_{ij}.

The marginal event rate of treatment t is $p_t = E[p_{kt}|\mu_t, \delta_t]$; for the logit link as in

Equation (2.1), p_t can be approximated as in Zeger et al., [44]

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

Using these marginal absolute risks of the T treatments, the risk ratio (RR) and risk difference (RD) for each pair of treatments can be estimated accordingly as $\text{RR}_{ij} = p_i/p_j$ and $\text{RD}_{ij} = p_i - p_j$. We can also compute two log odds ratio estimands: 1) the marginal log odds ratio between treatments i and j $\text{mLOR}_{ij} = \log \frac{p_i(1-p_j)}{p_j(1-p_i)}$, and 2) the conditional log odds ratio $\text{cLOR}_{ij} = \mu_i - \mu_j$, which is more commonly used in the meta-analyses literature. Agresti [45] (pp. 496–497) provided further details differences between these two LORs. Other link functions could also be used in Equation (2.1). For example, using the probit link $\Phi^{-1}(p_{kt}) = \mu_t + \nu_{kt}$ as in Zhang et al. [7], the marginal absolute risk has an exact form, $p_t = \Phi\left(\mu_t/\sqrt{1+\delta_t^2}\right)$, where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the cumulative distribution function of the standard normal distribution and its inverse, respectively. When the logit link is used, both cLOR and mLOR can be estimated; using the probit link, only mLOR can be directly estimated.

2.3 Prior specifications for the covariance matrix

This section describes specifications of prior distributions for the AB-NMA in Bayesian analysis. Specifically, we place vague $N(0, 100^2)$ priors on μ_t $(t = 1, \ldots, T)$, and discuss prior distributions for the covariance matrix $\mathbf{\Sigma}$. Two common ways to specify the prior distribution for the covariance matrix $\mathbf{\Sigma}$ are the natural conjugate prior for the multivariate normal likelihood, which treats the covariance matrix as a whole, and the separation strategy proposed by Barnard et al. [17], which decomposes the covariance matrix into separate parts as $\mathbf{\Sigma} = \mathbf{\Delta P \Delta}$ and assigns priors to the components $\mathbf{\Delta}$ and $\mathbf{P}$ separately. As above, $\mathbf{\Delta}$ is a diagonal matrix with standard deviation δ_i as its i^{th} diagonal element, and $\mathbf{P}$ is a correlation matrix with diagonal elements 1 and off-diagonal elements ρ_{ij}. The following subsections give more details about these two methods and their variations.

2.3.1 The Inverse-Wishart prior

The inverse-Wishart (IW) distribution for the covariance $\mathbf{\Sigma}$ is a conjugate prior for the multivariate normal likelihood, which can speed up computation compared to other priors. The density function of the IW distribution with degrees of freedom m ($> T-1$) and positive definite scale matrix $\mathbf{\Psi}$ is:

$$\pi(\mathbf{\Sigma}; \mathbf{\Psi}, m) \propto |\mathbf{\Psi}|^{-m/2} |\mathbf{\Sigma}|^{-(m+T+1)/2} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{\Psi})\right), \tag{2.2}$$

where $|\cdot|$ and $\text{tr}(\cdot)$ denote the determinant and trace of a matrix, respectively. The scale matrix $\mathbf{\Psi}$ is often selected to be the $T \times T$ identity matrix $\mathbf{I}$.

We now give some remarks about the IW prior's properties and problems. The marginal distribution of each δ_i^2 (the i^{th} diagonal element of $\mathbf{\Sigma}$) is the inverse-gamma distribution $IG(\frac{m-T+1}{2}, \frac{\psi_{ii}}{2})$ for $i = 1, \ldots, T$, where $\frac{m-T+1}{2}$ is the shape parameter and $\frac{\psi_{ii}}{2}$ is the scale parameter. Here, ψ_{ii} is the i^{th} diagonal element of $\mathbf{\Psi}$.

The IW prior has limitations. For example, simulation studies by Alvarez et al. [20] showed that although the estimated covariances (posterior means) Σ_{ij} $(i \neq j)$ are unbiased, the estimated correlations (posterior means) ρ_{ij} are biased towards zero and the estimated variance (posterior mean) Σ_{ii} is biased upward when the true variance Σ_{ii} is small. These biases are caused by the lack of prior density near zero in the marginal prior for the variance and the dependence induced by the IW prior between the correlation matrix $\mathbf{P}$ and the variances. Section 2.4's simulation studies also illustrate this problem.

Finally, the IW prior does not allow a user to specify different amounts of prior knowledge about different variance components; the single parameter m controls this uncertainty for all diagonal elements. To add more flexibility, the scaled IW prior [41] and the hierarchical half-t prior [46] have been proposed; however, such flexibility may be limited compared to the separation strategy [20].

2.3.2 The separation strategy

As mentioned, the separation strategy allows more flexibility by decomposing the covariance matrix $\mathbf{\Sigma}$ as $\mathbf{\Delta P \Delta}$ and placing independent prior distributions on the standard deviations δ_i and the correlation matrix $\mathbf{P}$. Popular priors for δ_i include the inverse-gamma prior for the variance (δ_i^2), the uniform prior between 0 and a certain upper

bound, the half-Cauchy [19] and the log-normal. [17] There is less consensus about choice of a prior for the correlation matrix $\mathbf{P}$, given the difficulty of constraining $\mathbf{P}$ to be positive definite. The following subsections elaborate this by discussing four possible choices.

The restricted inverse-Wishart prior and restricted Wishart Prior

We start with the restricted inverse-Wishart (RIW) prior mentioned by Barnard et al. [17], where the correlation matrix $\mathbf{P}$ follows an IW distribution with the restriction that its diagonal elements are fixed as 1. Specifically, let $\mathbf{Q} \sim IW_T(\mathbf{I}, m)$, then $\mathbf{P} = \mathbf{\Delta Q \Delta}$ follows a $RIW_T(m)$ distribution, where $\mathbf{\Delta}$ is a diagonal matrix with i^{th} diagonal element $Q_{ii}^{-1/2}$ and Q_{ii} is the i^{th} diagonal element of $\mathbf{Q}$. In the RIW prior, the ρ_{ij} have the same marginal distributions for all $i \neq j$:

$$\pi(\rho_{ij}) \propto (1-\rho_{ij}^2)^{\frac{m-T-1}{2}}, \quad -1 < \rho_{ij} < 1, \tag{2.3}$$

a beta distribution, $Beta(\frac{m-T+1}{2}, \frac{m-T+1}{2})$, on the interval $[-1, 1]$, which is uniform if $m = T+1$. Hence, the $RIW_T(T+1)$ prior implies marginally uniform-distributed correlation coefficients. This prior is distinguished from the jointly uniform prior $\pi(\mathbf{P}) \propto 1$, which is a special case of the LKJ prior [43]. Actually the LKJ prior is equivalent, in a specific sense, to the restricted Wishart (RW) prior, which we now discuss. (Appendix B.1 gives the proof.)

To define the RW prior, as for the RIW prior let $\mathbf{Q}^*$ follow a Wishart distribution $W_T(\mathbf{I}, m^*)$, then $\mathbf{P} = \mathbf{\Delta Q^* \Delta}$ follows a restricted Wishart distribution $RW_T(m^*)$ with degrees of freedom m^* ($> T-1$); again, $\mathbf{\Delta}$ is a diagonal matrix with i^{th} diagonal element $1/\sqrt{Q_{ii}^*}$. The RW prior has density $\pi(\mathbf{P}) \propto |\mathbf{P}|^{(m^*-T-1)/2}$; if $m^* = T+1$, this is uniform on a compact subspace of the $T(T-1)/2$ dimensional hypercube $(-1, 1)^{T(T-1)/2}$. Also, the ρ_{ij} for all $i \neq j$ follow the same beta distribution $Beta(\frac{m^*-1}{2}, \frac{m^*-1}{2})$ on $[-1, 1]$, which is $Beta(\frac{T}{2}, \frac{T}{2})$ if $m^* = T+1$. Based on this, the marginal distributions of the correlation coefficients ρ_{ij} in the jointly uniform prior tend to place density close to zero as the dimension T increases, which is the key difference from the marginally uniform prior. Furthermore, the conditional distribution of the correlation coefficient ρ_{ij} given ρ_{rs} with $(i, j) \neq (r, s)$ in the RW prior also differs from that in the RIW prior.

Although the closed form for the conditional distribution of ρ_{ij} given ρ_{rs} only exists

for special cases, we can visualize these conditional distributions as in Figure 2.1, which compares the RIW and RW distributions when the dimension T is 3 or 10. As shown in Figure 2.1, the RIW distribution puts a marginally uniform prior on each ρ_{ij} while the marginal prior on the individual correlation ρ_{12} under the RW distribution (which is jointly uniform) concentrates around zero, more so as the dimension T increases. Figure 2.1 also shows that compared to the RW distribution, the RIW distribution puts more density close to 1 for ρ_{12} when both ρ_{23} and ρ_{13} are larger than 0.6. Interestingly, of 1,000,000 random draws from the RW distribution with $T = 10$, only two satisfied the condition that both ρ_{23} and ρ_{13} are larger than 0.8. By contrast, randomly drawing correlation matrices from the RIW distribution, 26,174 out of 1,000,000 satisfied this condition. The above findings may have implications for prior choice in applications. For example, in NMA, we might prefer the RIW prior to the RW prior because the RW prior places less density on all correlation elements ρ_{ij} being large.

Exchangeable correlation structure

To reduce model complexity, Lin et al. [47] proposed an exchangeable structure (EQ prior) for $\mathbf{P}$, where all off-diagonal elements ρ_{ij} are assumed equal to a common value ρ. To keep $\boldsymbol{P}$ positive definite, ρ must be larger than $-\frac{1}{T-1}$, so we may specify a vague uniform prior for ρ on $(-\frac{1}{T-1}, 1)$.

2.4 Simulation studies

We conducted simulation studies to compare the performance of AB-NMA using different priors (IW, RW, RIW and EQ) in terms of bias and coverage probability, under different mechanisms for selecting treatment arms to be included in each study of the AB-NMA.

2.4.1 Simulation settings

We describe the simulation studies using three main steps: first, how we generated a complete data set; second, how we applied the missing data mechanisms to omit some arms from studies; and third, which estimands and priors we chose. In the simulation,

we fit the AB-NMA with different priors to each simulated dataset, saved estimates of the estimands, and described the performance measures.

Each simulated NMA dataset $\{D_1, \ldots, D_K\}$ had binary outcome data $D_k = \{(r_{kt}, n_{kt}), t = 1, 2, 3\}$ from 18 studies ($K = 18$) comparing three treatments ($T = 3$), denoted 1, 2, and 3. The number of patients n_{kt} in each arm was fixed at 500. The number of simulated datasets in each setting was 1000. First, we generated complete datasets under the AB model specification in Equation (2.1) using the logit link and setting $(\mu_1, \mu_2, \mu_3) = (\mu_2 + 0.5, \mu_2, \mu_2 - 0.5)$, $(\nu_{k1}, \nu_{k2}, \nu_{k3})' \sim MVN(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma} = \mathbf{\Delta P \Delta}$ with standard deviations $(\delta_1, \delta_2, \delta_3)$ and correlation matrix $\mathbf{P}$ with entries $(\rho_{12}, \rho_{13}, \rho_{23})$. In Scenario I, we chose $(\delta_1, \delta_2, \delta_3) = (0.7, 0.4, 0.1)$, $(\mu_1, \mu_2, \mu_3) = (-1.0, -1.5, -2.0)$ and $(\rho_{12}, \rho_{13}, \rho_{23}) = (\frac{2}{3}, \frac{4}{9}, \frac{2}{3})$. We also conducted additional simulations and analyses to confirm our findings under different correlation structures and μ's. For these, we set $(\delta_1, \delta_2, \delta_3) = (0.7, 0.4, 0.1)$ and considered three different values of $\mu_2 \in \{-0.5, -1.5, -2.5\}$ with $\mu_1 = \mu_2 + 0.5$ and $\mu_3 = \mu_2 - 0.5$. We also considered three choices for $\mathbf{P}$: high correlation scenario with a common between-treatment correlation $\rho = 0.7$, low correlation scenario with $\rho = 0.2$, and mixed between-treatment correlation scenario with $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.7, 0.1, 0.4)$.

After generating the complete dataset, we omitted treatment arms to create partially missing data under MAR and MNAR with respect to absolute effects. For each missingness setting, we obtained two different sets of nine two-arm studies: one set compared treatments 1 and 2, and another compared treatments 2 and 3. To generate partially missing data under the MAR assumption, we first kept all treatment 2 data observed and ranked the studies in ascending order by r_{k2}/n_{k2}. Then, we made treatment 1 missing in the first nine studies in this ordering and treatment 3 missing in the last nine studies. This created datasets in which the results for the control arm (treatment 2) had improved over time (r_{k2}/n_{k2} increased with k), while only studies with treatment 1 (more advanced regimen) versus 2 were available in the more recent period and studies with treatment 3 (less advanced regimen) versus 2 were available in the earlier period. Next, we used the following strategy to create missing data under MNAR with respect to absolute effects. We used m_{kt} ($k = 1, \ldots, K$ and $t = 1, \ldots, T$) to indicate missingness of the t^{th} treatment in the k^{th} study; $m_{kt} = 1$ indicated missing

and $m_{kt} = 0$ indicated not missing. First, we assumed all treatment 2 data were observed, so $m_{k2} = 0$, $k = 1, \ldots, K$. Then we determined the missingness of treatment 1 based on the data for all 3 treatments. After determining m_{k1}, the missingness of treatment 3 followed automatically: $m_{k3} = 1 - m_{k1}$. The model to generate the missingness indicators was:

$$\begin{aligned} &m_{k1} \sim \text{Bernoulli}(\pi_{k1}), k = 1, \ldots, K; \\ &\text{logit}(\pi_{k1}) = \beta_0 + \beta_1 \times [\text{logit}(r_{k1}/n_{k1}) + \text{logit}(r_{k2}/n_{k2}) + \text{logit}(r_{k3}/n_{k3})], \end{aligned} \tag{2.4}$$

where π_{k1} is the probability of treatment 1 being missing in study k. The parameters β_0 and β_1 were pre-defined to control the average number of studies without treatment 1 to be 9 in each scenario: $\beta_1 = 1$ and $\beta_0 = -\mu_1 - \mu_2 - \mu_3$. A continuity correction of 0.5 was applied to both r and n when r_{ki} was zero.

We focus on these estimands: the two kinds of log odds ratio comparing treatments i and j, $\text{mLOR}_{ij} = \log(\frac{p_i(1-p_j)}{p_j(1-p_i)})$ and $\text{cLOR}_{ij} = \mu_i - \mu_j$, the absolute risk of an event for treatment t (p_t), the standard deviation of log-odds for treatment t (δ_t), and the correlation between treatment-specific (treatments i and j) log-odds (ρ_{ij}). When applying the AB-IW model, we set the prior for the covariance matrix $\mathbf{\Sigma}$ to be $IW_T(\mathbf{I}, T+1)$. For models using the RW, RIW and EQ priors, we imposed independent uniform priors $U(0, 5)$ on standard deviations δ_i, $i = 1, \ldots, T$, then set the $RW_T(T+1)$, $RIW_T(T+1)$ prior for the correlation matrix $\mathbf{P}$ and the uniform prior $U(-\frac{1}{T-1}, 1)$ for the correlation coefficient ρ, respectively. Clearly, the true standard deviations $(\delta_1, \delta_2, \delta_3) = (0.7, 0.4, 0.1)$ were not close to the center of the prior U(0,5), nor were the true correlations $(\rho_{12}, \rho_{13}, \rho_{23}) = (\frac{2}{3}, \frac{4}{9}, \frac{2}{3})$ close to the center of the RW, RIW, and EQ priors.

We implemented the models using Stan [48] in conjunction with R [49]. We chose posterior mean and 95% equal tailed credible interval as point and interval estimates respectively. To measure the performance of different methods, we used bias and coverage probability of 95% credible interval.

2.4.2 Simulation results

Table 2.1 summarizes bias of the estimates and coverage probability (CP) of the 95% credible intervals given by four AB models (AB-IW, AB-RW, AB-RIW and AB-EQ)

under different missingness settings (no missing, MAR, and MNAR). The table includes the log odds ratio comparing treatments i and j, mLOR_{ij} and cLOR_{ij}, the absolute risk of an event for treatment t (p_t), the standard deviation of log-odds for treatment t (δ_t), and the correlation between treatment-specific (treatments i and j) log-odds (ρ_{ij}).

All four AB models gave unbiased estimates and good coverage probabilities for the log odds ratios mLOR_{ij} and cLOR_{ij} and the absolute risks p_t for complete datasets ("no missing", i.e., no omitted arms). However, the AB-IW model produced biased estimates of the standard deviations δ_t and correlations ρ_{ij} and intervals with low coverage probabilities. In particular, the AB-IW method had bias 0.202 for δ_3, which was large relative to the true value $\delta_3 = 0.1$, while all AB-RW, AB-RIW and AB-EQ had almost no bias (0.005). Also, for the IW prior the bias of ρ_{23} (-0.511) was so large that this method estimated little posterior association between treatments 2 and 3, although the true correlation was 0.667. In summary, the IW prior gave upwardly biased estimates for small variances and correlation estimates biased towards zero, especially when the variance was small. In contrast, the AB-RW, AB-RIW and AB-EQ priors gave estimates of standard deviations and correlations with much smaller biases (albeit not exactly zero) with good coverage probability.

For complete datasets, mis-estimating correlations between treatment-specific log-odds did not bias estimates of log odds ratios and absolute risks. However, in datasets with missing entries (arms omitted under MAR/MNAR), underestimation of correlations, especially for the star-shaped network structure used here, gave estimated relative effects with larger bias and worse coverage probability. Underestimation of correlations increased under MAR/MNAR compared to complete data for all priors considered; e.g., for AB-IW, the estimated bias for $(\rho_{12}, \rho_{13}, \rho_{23})$ was $(-0.17, -0.32, -0.51)$, $(-0.45, -0.43, -0.62)$ and $(-0.36, -0.41, -0.59)$ in the no missing, MAR, and MNAR scenarios, respectively. Also, while the AB-RIW and AB-EQ priors gave similar results under MNAR for mLOR_{ij} and cLOR_{ij}, the extra but reasonable assumption of equal correlations between treatment-specific log-odds reduced the bias of estimated correlations and log odds ratios under MAR. The performance of AB-RW prior is slightly worse than AB-RIW prior in terms of estimated relative effects and correlations because RW prior places less density on all correlation elements being large, as was mentioned regarding Figure 2.1.

We verified our findings using additional simulations, shown in Table 2.2, under various μ's and correlation structures. The AB-RW prior gave smaller bias and higher coverage probability than the AB-IW prior for mLOR_{13} (cLOR_{13}), the AB-RIW prior gave smaller bias and higher coverage probability than the AB-RW prior for mLOR_{13} (cLOR_{13}), and the AB-EQ prior gave much smaller bias than the AB-RIW prior for mLOR_{13} (cLOR_{13}) under MAR/MNAR. Also, AB-RIW and AB-EQ generally provided more reliable estimates of absolute risks (p_1) than AB-IW. These results were in line with our findings in Table 2.1.

2.5 Case studies

One pitfall of AB-NMA is that the data may not provide enough information about the variances of log-odds for some treatments or about correlations between treatment-specific log-odds. In a Bayesian analysis, such a lack of information may cause the posterior to be dominated by prior information. This section uses three examples with different levels of information to examine the performance of different AB methods. Here, we focus more on posterior distributions of standard deviations than on correlations for two reasons. First, Figure 2.1 shows, for the RIW prior, that the positive definiteness property already imposes strong prior information on relationships among individual correlations and the IW and EQ priors make more strict assumptions than the RIW prior, as shown in Sections 2.3.1 and 2.3.2, respectively. Second, NMAs commonly provide little information about correlations and no practical remedy is available to avoid the influence of the prior on the posterior. We considered seven models in these case studies: AB-IW, AB-RW, AB-RIW, AB-EQ, AB-RW-EV, AB-RIW-EV and AB-EQ-EV. The latter three new models were based on AB-RW, AB-RIW and AB-EQ with the further assumption that all δ_i ($i = 1, \ldots, T$) were equal to δ, with a uniform prior $U(0, 5)$.

2.5.1 Example 1: smoking abstinence data

Mills et al. [50] summarized results of 101 trials with 31,321 individuals comparing 4 interventions for the primary outcome of abstinence from smoking at least 4 weeks post-target quit date. Figure 2.2A shows the network plot of the data with treatments

1–4 being control, nicotine replacement therapy (NRT), bupropion, and varenicline, respectively. We used the AB-IW, AB-RW, AB-RIW, and AB-EQ models to analyze this dataset.

Table 2.3 summarizes results for marginal log odds ratios, mLOR_{ij}, comparing treatments i and j, the absolute risk of treatment t (p_t), the standard deviation of log-odds for treatment t (δ_t), the correlation between treatment-specific (treatments i and j) log-odds (ρ_{ij}), the probability that treatment t ranks i^{th} (Rank_{ti}; a higher rank means a larger proportion with events), and the deviance information criterion (DIC) [51] using the different models. The four AB methods (AB-IW, AB-RW, AB-RIW, and AB-EQ) gave similar results except for the log odds ratio of bupropion versus varenicline (mLOR_{34}) and NRT versus varenicline (mLOR_{24}). Although the AB-RIW and AB-EQ models gave slightly different posterior means of the log odds ratios, they led to the same conclusions: all active therapies (NRT, bupropion, and varenicline) increased smoking abstinence compared to control in the short term, and varenicline was more effective than NRT or bupropion. The rank probabilities also indicated that varenicline had the best results ($\text{Rank}_{41} > 0.9$) and control had the worst ($\text{Rank}_{14} = 1$). Because this dataset was large and provided enough information — 101 trials comparing four treatments, each included in at least 9 trials (varenicline was in 9 trials) — it is not surprising that these AB methods gave similar conclusions.

2.5.2 Example 2: serious vascular events prevention data

This dataset, reported by Thijs et al. [52], consisted of 24 antiplatelet trials involving 42,688 patients after transient ischaemic attack (TIA) or stroke and compared 5 regimens: 1) placebo, 2) aspirin (ASA) plus thienopyridines (THIENO), 3) aspirin, 4) aspirin plus dipyridamole (DP), and 5) thienopyridines. The outcome was occurrence of a serious vascular event, including myocardial infarction and vascular death after TIA or stroke. Figure 2.2B shows the network plot; ASA plus thienopyridines was included in only 3 trials. As we had limited information about the variance of the effect of ASA plus THIENO, one may wonder how its posterior distribution was influenced by its prior. Hence, we added three models: 1) the AB-RW model assuming equal variances (δ_t^2) (AB-RW-EV), 2) the AB-RIW model assuming equal variances (δ_t^2) (AB-RIW-EV), and 3) the AB-EQ model assuming equal variances (AB-EQ-EV).

Table 2.4 summarizes the results. For the absolute proportion of serious vascular events for ASA plus THIENO (p_2), which had limited information, both the AB-RIW and AB-EQ methods gave an unrealistically high upper bound for the 95% credible interval of δ_2 (RIW: 0.19 to 3.67; EQ: 0.23 to 2.73). This resulted in wide credible intervals for the absolute risk p_2 (RIW: 0.06 to 0.53; EQ: 0.11 to 0.48) and potentially biased estimates. Such wide intervals meant that we could not find any significant relative effects involving ASA plus THIENO; for example, using AB-RIW, $\text{mLOR}_{12} = 0.24$ with 95% CI $(-1.52, 1.35)$. However, assuming equal variances (the AB-RIW-EV and AB-EQ-EV methods) narrowed the credible intervals of p_2 and δ_2. Using the AB-IW model, standard deviations δ_1 and δ_4 were potentially overestimated when comparing with the other two separation strategy priors (RIW and EQ), which caused disruptions in estimating absolute risks and relative risks related to the placebo and aspirin plus DP arms. Moreover, correlations ρ_{12}, ρ_{23}, ρ_{24}, and ρ_{25} were potentially underestimated by both AB-IW and AB-RW-EV models because of the lack of comparisons between these treatments. Although the AB-IW and AB-RW-EV models could also control the credible intervals of δ_t and p_t, DIC indicated that we may prefer AB-RIW-EV/AB-EQ-EV to AB-IW/AB-RW-EV (DIC: 74.91/74.55 vs. 89.97/87.08) and the DIC differences here were large enough to be of practical importance (larger than 5 units).

2.5.3 Example 3: postpartum haemorrhage prevention data

Hofmeyr et al. [53] reviewed 25 studies on prevention of postpartum haemorrhage (blood loss $\geq$ 1000 ml). This NMA compared four regimens: 1) other uterotonics, 2) misoprostol 600–800 mcg, 3) misoprostol 400–500 mcg, and 4) misoprostol <400 mcg as in Figure 2.2C. Table 2.5 summarizes the results; we focus on δ_4. As Figure 2.2C shows, only one trial directly compared treatment 4 with other treatments; therefore, for the AB-EQ method, δ_4's posterior (mean 2.48; 95% CI, 0.19, 4.85) was dominated by the $U(0,5)$ prior. However, this did not imply that AB-EQ was inferior to AB-EQ-EV. When confronted with a low-information situation, both the AB-EQ and AB-EQ-EV (or the AB-RIW and AB-RIW-EV) methods were useful: AB-EQ alerted us to a lack of information, while AB-EQ-EV showed the consequences of an extra assumption.

2.6 Summary and discussion

This chapter evaluated different prior distributions for the covariance matrix in an AB-NMA, including the IW prior and a separation strategy with uniform priors on standard deviations and the RW, RIW or EQ prior on correlations. We compared their performance using extensive simulation studies with data generated under different mechanisms for selecting each study's treatments. Separation strategies with a RIW or EQ prior on correlations performed much better than the IW prior in all situations in terms of bias and coverage probability of log odds ratios, absolute risks, variances and correlations. The commonly used IW prior often overestimated variances and underestimated correlations, which can lead to substantial bias for log odds ratios and absolute effects, especially under MNAR. The separation strategy with the EQ prior gave relatively small biases for log odds ratios under all conditions considered. These findings suggest that the separation strategy with the equal correlation prior is a much better choice of default vague prior in AB-NMA than the widely-used IW prior.

We conducted three case studies and compared separation strategies to the IW prior in terms of DIC. In the meta-analysis of serious vascular event prevention, the RIW and EQ priors had noticeably improved DIC compared to the IW and RW priors. However, estimating treatment-specific variances of log-odds in AB-NMA can be encumbered due to lack of information (i.e., most treatments are included in only a few trials). Here we proposed a straightforward but perhaps overly simple solution of assuming the treatment-specific variances are equal. This equal-variances assumption gave narrower credible intervals for treatment effects in the NMAs of serious vascular event prevention and postpartum haemorrhage prevention, and improved model fits. Thus, we suggest that NMA users consider sensitivity analyses with different assumptions on variance (homogeneous and heterogeneous variance assumptions) when using the separation strategy with the equal correlation prior.

Due to space limitation, we compared only four covariance priors. Other potential choices include the Cholesky decomposition [42] and the spherical decomposition [54]. These methods further decompose the correlation matrix $\mathbf{P}$ as $\mathbf{L}'\mathbf{L}$, where $\mathbf{L}$ is a $T \times T$ upper-triangular matrix. Then we can place a weakly informative prior on $\mathbf{L}$ through

a spherical parameterization. Technical details and applications to multivariate meta-analyses are in Lu and Ades [6], Wei and Higgins [55], and Lin et al. [56]. However, using this approach, the marginal distributions of the ρ_{ij} depend strongly on the indexes i and j; see, e.g., Figure 3 in Wei and Higgins [55]. The different marginal distributions lack statistical or clinical interpretations and meta-analysts may reasonably be concerned about the potential impact of using different marginal priors for different correlations.

We evaluated the performance of a Bayesian analysis with different prior specifications with simulations from a frequentist perspective (e.g., bias and coverage probability); one may argue that this may be philosophically inappropriate [57]. Nevertheless, this does not affect the goal of this chapter, i.e., providing some practical recommendations of prior specifications for the Bayesian AB-NMA with less bias and better coverage probability.

In the continuing debate [13,14] between proponents of the AB and CB approaches, White et al. [40] recently compared both approaches and concluded that 'both AB and CB models are suitable for the analysis of NMA data, but using random study intercepts requires a strong rationale such as relating treatment effects to study intercepts'. Perhaps such a rationale exists: Houwelingen et al. [58,59] pointed out that the assumption of exchangeable absolute treatment risks is reasonable in most meta-analyses, while Béliveau et al. [15] also mentioned that disconnected networks could benefit from being analyzed using random study intercepts. But perhaps this assumption appeared to be important only by accident, because of the choice of prior distributions: in our simulation study (Table 2.1), the separation strategy approach (AB-EV), compared with AB-IW, substantially reduced bias and potentially reduced the risk of using random study intercepts by improving estimation of the correlations between different treatments' random effects of log-odds.

While the most important variance parameter in a CB-NMA is the heterogeneity of the treatment contrasts [40], in the AB approach the variances of log-odds and correlations between treatment-specific log-odds are crucial, because: 1) variances of log-odds are needed to derive absolute risks and a much wider range of estimands, which is a key advantage of the AB approach; 2) correlations are critical to keep contrasts (relationships) between treatments stable, which can reduce the risk of assuming random study intercepts. Also, the AB-EQ-EV prior discussed in Section 2.5, with homogeneous

variance of treatment specific log-odds and homogeneous correlations, is equivalent to assuming homogeneous variance of treatment contrasts in the CB approach [10], in terms of the covariance structure.

Several other issues deserve further exploration and discussion. First, the uniform prior on standard deviations δ_i, $i = 1, \ldots, T$ in the AB model may cause upward bias when the true standard deviation is low. Second, if certain treatments are included in only a few trials in a NMA, the posteriors of their standard deviations may be dominated by their priors, which could lead to wide credible intervals for both the standard deviations and absolute risks, and thus bias the estimated absolute risks. In some NMAs, the assumption of equal variances might be too strong. Alternatively, one may model the standard deviations of study-specific log-odds of different treatments as random draws from a common distribution and thus allow borrowing strength in the estimation to shrink them in a data-dependent manner. The idea of extrapolation [37,60] could also be applied to incorporate external evidence about standard deviations. Third, since the posterior distributions of some parameters (e.g., correlation coefficients ρ_{ij} and marginal absolute risks p_i) could be skewed, posterior medians might be better summaries than posterior means. Finally, it seems that all priors considered here may systematically underestimate the correlations. Further research on alternative priors may be fruitful.

Figure 2.1: Conditional prior densities for ρ_{12} with dimension $T = 3$ (right panels) and dimension $T = 10$ (left panels). Each distribution is estimated from 1,000,000 random draws from the respective distribution (restricted inverse-Wishart in the first row and restricted Wishart in the second row), using the R function `geom_density()` with default settings to compute and draw kernel density estimates. Different colors denote different conditioning criteria; e.g., the purple line is the conditional distribution of ρ_{12} given ρ_{23} and ρ_{13} both larger than 0.8. We did not draw the purple density for the restricted Wishart with dimension 10 because only 2 out of 1,000,000 random draws had ρ_{23} and ρ_{13} both greater than 0.8.

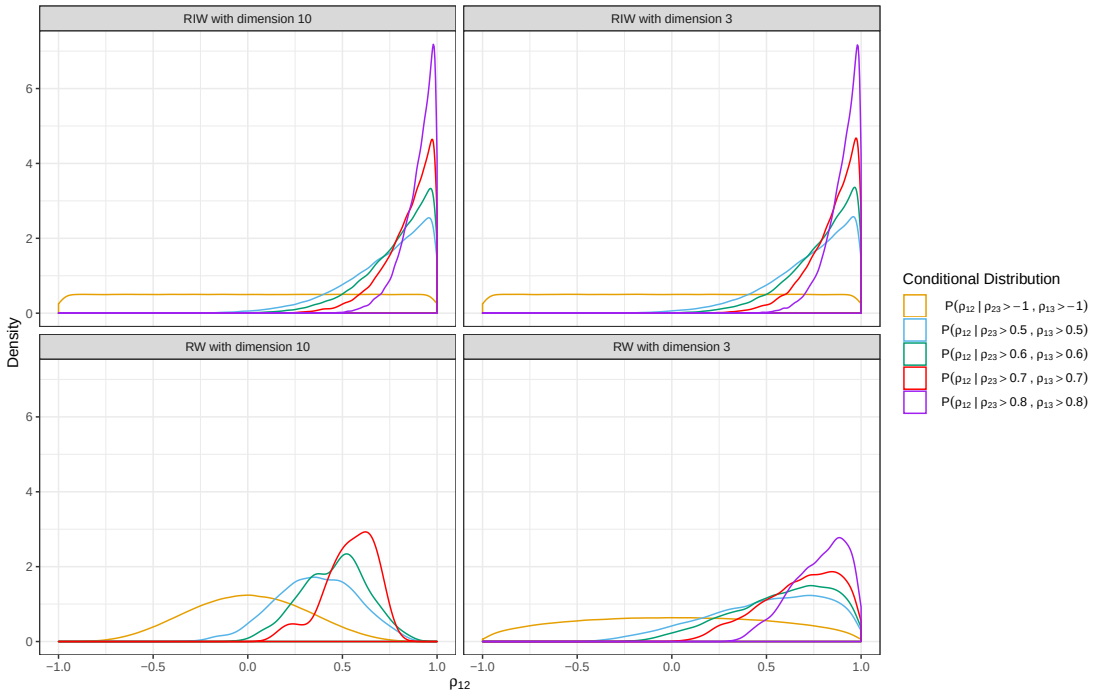


Table 2.1: Simulation results comparing data generated under Scenario I with $(\delta_1, \delta_2, \delta_3) = (0.7, 0.4, 0.1)$, $(\mu_1, \mu_2, \mu_3) = (-1.0, -1.5, -2.0)$ and $(\rho_{12}, \rho_{13}, \rho_{23}) = (\frac{2}{3}, \frac{4}{9}, \frac{2}{3})$. Performance of AB models (IW, RW, RIW, and EQ) under different missingness mechanisms (No missing, MAR, MNAR) is shown: the bias of estimates (posterior mean) of cLOR_{ij}, mLOR_{ij}, p_t, δ_t, and ρ_{ij} and coverage probability of 95% credible intervals. MCMC errors are at the 0.01 level.

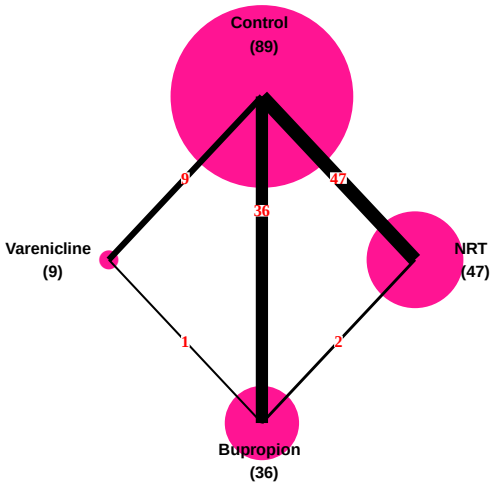
		Bias (Coverage Probability)											
Parameter	Truth	AB-IW	AB-RW	AB-RIW	AB-EQ	AB-IW	AB-RW	AB-RIW	AB-EQ	AB-IW	AB-RW	AB-RIW	AB-EQ
Scenario I													
Missing Setting		**No missing**				**MAR**				**MNAR**			
cLOR_{12}	0.50	-0.001(0.976)	-0.003(0.970)	-0.003(0.962)	-0.003(0.964)	0.243(0.900)	0.136(0.954)	0.087(0.964)	0.032(0.962)	-0.226(0.884)	-0.186(0.921)	-0.172(0.918)	-0.151(0.916)
cLOR_{13}	1.00	0.010(0.957)	0.001(0.950)	0.000(0.946)	0.001(0.937)	0.293(0.904)	0.174(0.946)	0.120(0.959)	0.034(0.970)	-0.240(0.927)	-0.205(0.921)	-0.188(0.927)	-0.155(0.937)
cLOR_{23}	0.50	0.011(0.990)	0.004(0.965)	0.004(0.961)	0.004(0.960)	0.050(1.000)	0.039(0.977)	0.033(0.980)	0.003(0.978)	-0.014(0.999)	-0.019(0.970)	-0.017(0.976)	-0.004(0.965)
mLOR_{12}	0.54	-0.010(0.979)	0.002(0.969)	0.002(0.965)	0.003(0.966)	0.208(0.913)	0.145(0.956)	0.112(0.966)	0.066(0.967)	-0.230(0.871)	-0.164(0.923)	-0.150(0.921)	-0.133(0.922)
mLOR_{13}	1.07	-0.012(0.962)	0.012(0.957)	0.014(0.958)	0.015(0.945)	0.221(0.950)	0.187(0.944)	0.148(0.961)	0.072(0.971)	-0.280(0.898)	-0.181(0.927)	-0.162(0.933)	-0.133(0.936)
mLOR_{23}	0.54	-0.002(0.994)	0.011(0.965)	0.012(0.964)	0.012(0.964)	0.013(1.000)	0.041(0.980)	0.036(0.980)	0.006(0.979)	-0.050(0.998)	-0.016(0.973)	-0.012(0.978)	-0.001(0.966)
p_1	0.28	0.003(0.940)	0.005(0.945)	0.005(0.951)	0.005(0.947)	0.051(0.855)	0.038(0.948)	0.032(0.959)	0.021(0.965)	-0.037(0.864)	-0.025(0.926)	-0.022(0.930)	-0.019(0.928)
p_2	0.19	0.003(0.977)	0.002(0.959)	0.003(0.962)	0.003(0.958)	0.003(0.974)	0.003(0.962)	0.003(0.962)	0.003(0.960)	0.003(0.975)	0.003(0.964)	0.003(0.964)	0.003(0.966)
p_3	0.12	0.002(1.000)	0.000(0.965)	0.000(0.965)	0.000(0.963)	0.002(1.000)	-0.003(0.976)	-0.002(0.974)	0.001(0.981)	0.008(1.000)	0.003(0.956)	0.003(0.960)	0.002(0.962)
δ_1	0.70	0.023(0.964)	0.052(0.948)	0.064(0.936)	0.067(0.934)	-0.044(0.980)	0.086(0.969)	0.131(0.957)	0.132(0.950)	-0.038(0.975)	0.063(0.957)	0.077(0.956)	0.079(0.958)
δ_2	0.40	0.068(0.932)	0.029(0.961)	0.037(0.957)	0.038(0.954)	0.066(0.938)	0.034(0.958)	0.039(0.953)	0.039(0.957)	0.067(0.934)	0.034(0.955)	0.040(0.950)	0.039(0.955)
δ_3	0.10	0.202(0.000)	0.004(0.980)	0.005(0.981)	0.005(0.979)	0.296(0.000)	0.035(0.977)	0.041(0.975)	0.041(0.977)	0.295(0.000)	0.031(0.988)	0.032(0.986)	0.032(0.982)
ρ_{12}	0.67	-0.170(0.927)	-0.100(0.961)	-0.067(0.960)	-0.060(0.955)	-0.454(0.754)	-0.330(0.955)	-0.270(0.971)	-0.180(0.982)	-0.359(0.786)	-0.247(0.939)	-0.200(0.948)	-0.146(0.959)
ρ_{13}	0.44	-0.316(0.914)	-0.196(0.986)	-0.091(0.989)	0.162(0.820)	-0.430(1.000)	-0.406(1.000)	-0.351(0.999)	0.042(0.970)	-0.411(1.000)	-0.363(1.000)	-0.280(1.000)	0.076(0.956)
ρ_{23}	0.67	-0.511(0.098)	-0.293(0.974)	-0.228(0.983)	-0.060(0.955)	-0.623(0.293)	-0.546(0.983)	-0.508(0.989)	-0.180(0.982)	-0.593(0.211)	-0.471(0.973)	-0.423(0.984)	-0.146(0.959)

Table 2.2: Additional simulations comparing performance (bias and coverage probability of 95% credible intervals) of AB-IW, AB-RIW and AB-EQ with respect to posterior mean of cLOR_{13}, mLOR_{13} and p_1 under MAR and MNAR using data generated with $(\mu_1, \mu_2, \mu_3) = (\mu_2 + 0.5, \mu_2, \mu_2 - 0.5)$, $(\delta_1, \delta_2, \delta_3) = (0.7, 0.4, 0.1)$, $\mu_2 \in \{-0.5, -1.5, -2.5\}$ and correlation matrix $\mathbf{P}$ chosen from low, high and mixed correlation scenarios. MCMC errors are at the 0.01 level.

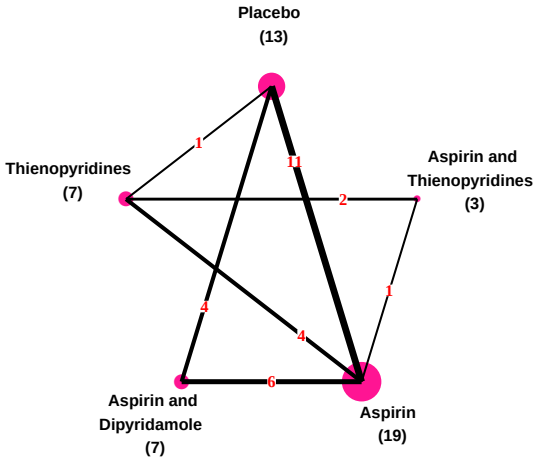
			cLOR_{13}: Bias (Coverage Probability)					mLOR_{13}: Bias (Coverage Probability)					p_1: Bias (Coverage Probability)				
Scenario	μ_2	Correlation structure	Truth	AB-IW	AB-RW	AB-RIW	AB-EQ	Truth	AB-IW	AB-RW	AB-RIW	AB-EQ	Truth	AB-IW	AB-RW	AB-RIW	AB-EQ
MAR																	
1	-0.50	High	1.00	0.302(0.883)	0.175(0.945)	0.117(0.957)	0.036(0.967)	1.00	0.263(0.913)	0.170(0.948)	0.119(0.959)	0.048(0.968)	0.50	0.059(0.847)	0.033(0.948)	0.023(0.954)	0.011(0.967)
2	-0.50	Low	1.00	0.090(0.979)	0.058(0.981)	0.045(0.978)	-0.063(0.978)	1.00	0.061(0.983)	0.053(0.980)	0.042(0.977)	-0.045(0.977)	0.50	0.017(0.961)	0.011(0.976)	0.009(0.971)	-0.009(0.976)
3	-0.50	Mixed	1.00	0.280(0.896)	0.152(0.946)	0.103(0.962)	0.042(0.961)	1.00	0.241(0.920)	0.147(0.945)	0.105(0.958)	0.049(0.961)	0.50	0.057(0.852)	0.031(0.949)	0.022(0.955)	0.016(0.953)
4	-1.50	High	1.00	0.291(0.895)	0.159(0.953)	0.099(0.965)	0.017(0.971)	1.07	0.218(0.939)	0.170(0.954)	0.126(0.962)	0.053(0.972)	0.28	0.051(0.858)	0.034(0.951)	0.027(0.962)	0.018(0.966)
5	-1.50	Low	1.00	0.093(0.982)	0.051(0.979)	0.035(0.977)	-0.069(0.977)	1.07	0.043(0.990)	0.087(0.980)	0.081(0.977)	-0.011(0.976)	0.28	0.020(0.965)	0.024(0.977)	0.024(0.977)	0.007(0.979)
6	-1.50	Mixed	1.00	0.280(0.903)	0.151(0.947)	0.097(0.961)	0.026(0.968)	1.07	0.207(0.946)	0.160(0.950)	0.121(0.960)	0.055(0.966)	0.28	0.052(0.856)	0.036(0.946)	0.029(0.955)	0.023(0.959)
7	-2.50	High	1.00	0.304(0.913)	0.181(0.946)	0.124(0.965)	0.031(0.973)	1.14	0.186(0.963)	0.195(0.949)	0.160(0.962)	0.074(0.981)	0.14	0.031(0.860)	0.027(0.933)	0.024(0.944)	0.018(0.955)
8	-2.50	Low	1.00	0.108(0.985)	0.059(0.983)	0.043(0.980)	-0.075(0.978)	1.14	0.026(0.995)	0.124(0.981)	0.123(0.975)	0.017(0.980)	0.14	0.015(0.969)	0.024(0.971)	0.025(0.968)	0.012(0.980)
9	-2.50	Mixed	1.00	0.289(0.919)	0.168(0.956)	0.115(0.966)	0.027(0.971)	1.14	0.172(0.977)	0.183(0.954)	0.151(0.966)	0.066(0.981)	0.14	0.031(0.870)	0.027(0.935)	0.024(0.948)	0.020(0.951)
MNAR																	
1	-0.50	High	1.00	-0.252(0.914)	-0.208(0.905)	-0.189(0.913)	-0.153(0.926)	1.00	-0.261(0.897)	-0.197(0.903)	-0.179(0.909)	-0.148(0.922)	0.50	-0.052(0.853)	-0.041(0.907)	-0.038(0.914)	-0.033(0.920)
2	-0.50	Low	1.00	-0.250(0.897)	-0.250(0.884)	-0.248(0.889)	-0.227(0.895)	1.00	-0.256(0.880)	-0.228(0.880)	-0.226(0.885)	-0.209(0.891)	0.50	-0.054(0.836)	-0.051(0.876)	-0.051(0.879)	-0.047(0.890)
3	-0.50	Mixed	1.00	-0.222(0.942)	-0.178(0.923)	-0.164(0.930)	-0.141(0.933)	1.00	-0.231(0.927)	-0.168(0.922)	-0.155(0.928)	-0.136(0.932)	0.50	-0.049(0.862)	-0.038(0.915)	-0.036(0.917)	-0.034(0.920)
4	-1.50	High	1.00	-0.244(0.917)	-0.207(0.905)	-0.188(0.906)	-0.155(0.925)	1.07	-0.285(0.870)	-0.185(0.910)	-0.165(0.913)	-0.137(0.925)	0.28	-0.037(0.860)	-0.025(0.916)	-0.022(0.916)	-0.019(0.921)
5	-1.50	Low	1.00	-0.239(0.904)	-0.243(0.896)	-0.241(0.894)	-0.223(0.902)	1.07	-0.274(0.871)	-0.198(0.906)	-0.195(0.897)	-0.182(0.912)	0.28	-0.037(0.846)	-0.029(0.912)	-0.028(0.914)	-0.027(0.923)
6	-1.50	Mixed	1.00	-0.204(0.937)	-0.170(0.923)	-0.154(0.932)	-0.130(0.937)	1.07	-0.247(0.903)	-0.148(0.921)	-0.132(0.934)	-0.111(0.936)	0.28	-0.033(0.885)	-0.021(0.928)	-0.018(0.934)	-0.017(0.930)
7	-2.50	High	1.00	-0.253(0.926)	-0.234(0.916)	-0.214(0.923)	-0.177(0.929)	1.14	-0.336(0.860)	-0.205(0.925)	-0.183(0.924)	-0.152(0.941)	0.14	-0.021(0.875)	-0.012(0.932)	-0.009(0.938)	-0.008(0.940)
8	-2.50	Low	1.00	-0.236(0.910)	-0.252(0.907)	-0.252(0.901)	-0.233(0.908)	1.14	-0.311(0.855)	-0.193(0.932)	-0.189(0.933)	-0.178(0.938)	0.14	-0.020(0.855)	-0.011(0.936)	-0.011(0.929)	-0.010(0.939)
9	-2.50	Mixed	1.00	-0.229(0.936)	-0.210(0.930)	-0.193(0.936)	-0.162(0.943)	1.14	-0.310(0.882)	-0.179(0.933)	-0.160(0.943)	-0.135(0.950)	0.14	-0.020(0.873)	-0.010(0.937)	-0.008(0.935)	-0.008(0.941)

Figure 2.2: Network plots of the three example datasets. Each node in a plot stands for a treatment and each edge represents a direct comparison between two treatments. Vertex size is proportional to the number (in parenthesis) of direct comparisons containing that treatment; edge thickness is proportional to the number (in red) of direct comparisons.

(A) Smoking abstinence



(B) Antiplatelet regimens



(C) PH prevention

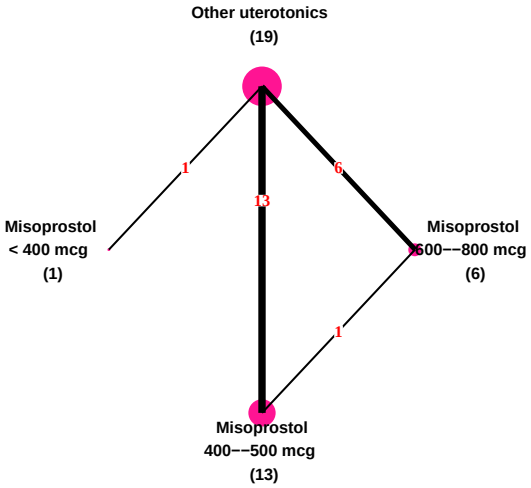


Table 2.3: Smoking abstinence data: comparison of posterior means and 95% credible intervals under 4 models, specifically mLOR_{ij} comparing the i^{th} and j^{th} treatment, absolute risk of events for the t^{th} treatment (p_t), standard deviation of log-odds for the t^{th} treatment (δ_t), correlation between treatment-specific (treatments i and j) log-odds (ρ_{ij}), and the i^{th} rank probability of the t^{th} treatment (Rank_{ti}). Regimen labels: (1) control, (2) nicotine replacement therapy, (3) bupropion, (4) varenicline.

	Point Estimate (95% Credible Interval)			
Parameter	AB-IW	AB-RW	AB-RIW	AB-EQ
mLOR_{12}	-0.58 (-0.74, -0.43)	-0.59 (-0.75, -0.42)	-0.58 (-0.73, -0.43)	-0.57 (-0.71, -0.43)
mLOR_{13}	-0.67 (-0.83, -0.51)	-0.68 (-0.83, -0.53)	-0.67 (-0.80, -0.53)	-0.66 (-0.81, -0.51)
mLOR_{14}	-0.90 (-1.22, -0.57)	-0.89 (-1.15, -0.59)	-0.94 (-1.16, -0.71)	-0.93 (-1.16, -0.73)
mLOR_{23}	-0.09 (-0.30, 0.12)	-0.09 (-0.30, 0.11)	-0.09 (-0.28, 0.11)	-0.09 (-0.28, 0.10)
mLOR_{24}	-0.32 (-0.66, 0.02)	-0.30 (-0.59, 0.01)	-0.36 (-0.61, -0.11)	-0.36 (-0.61, -0.13)
mLOR_{34}	-0.23 (-0.57, 0.11)	-0.20 (-0.49, 0.10)	-0.27 (-0.51, -0.03)	-0.27 (-0.52, -0.04)
Rank_{11}	0.00	0.00	0.00	0.00
Rank_{21}	0.01	0.01	0.00	0.00
Rank_{31}	0.08	0.08	0.01	0.01
Rank_{41}	0.90	0.91	0.98	0.99
p_1	0.28 (0.24, 0.32)	0.28 (0.24, 0.31)	0.28 (0.24, 0.31)	0.28 (0.24, 0.31)
p_2	0.41 (0.36, 0.45)	0.41 (0.36, 0.45)	0.41 (0.36, 0.45)	0.41 (0.36, 0.45)
p_3	0.43 (0.38, 0.48)	0.43 (0.39, 0.48)	0.43 (0.38, 0.47)	0.43 (0.38, 0.47)
p_4	0.49 (0.41, 0.56)	0.48 (0.42, 0.55)	0.50 (0.44, 0.55)	0.50 (0.45, 0.55)
δ	.	.	.	.
δ_1	0.89 (0.76, 1.06)	0.88 (0.74, 1.04)	0.91 (0.77, 1.08)	0.91 (0.77, 1.08)
δ_2	0.83 (0.66, 1.03)	0.82 (0.66, 1.02)	0.85 (0.68, 1.06)	0.87 (0.69, 1.08)
δ_3	0.79 (0.64, 0.97)	0.76 (0.62, 0.94)	0.78 (0.63, 0.96)	0.78 (0.63, 0.97)
δ_4	0.54 (0.32, 0.90)	0.42 (0.23, 0.75)	0.49 (0.27, 0.82)	0.48 (0.28, 0.81)
ρ	.	.	.	0.87 (0.78, 0.93)
ρ_{12}	0.80 (0.65, 0.89)	0.78 (0.62, 0.89)	0.83 (0.69, 0.92)	.
ρ_{13}	0.85 (0.73, 0.93)	0.89 (0.74, 0.97)	0.92 (0.81, 0.98)	.
ρ_{14}	0.57 (-0.15, 0.89)	0.60 (-0.14, 0.94)	0.86 (0.48, 0.98)	.
ρ_{23}	0.76 (0.51, 0.90)	0.68 (0.32, 0.94)	0.83 (0.57, 0.98)	.
ρ_{24}	0.51 (-0.21, 0.86)	0.47 (-0.31, 0.92)	0.79 (0.33, 0.98)	.
ρ_{34}	0.53 (-0.18, 0.87)	0.53 (-0.25, 0.95)	0.86 (0.42, 0.99)	.
DIC	331.38	332.70	332.40	330.30
$\bar{D}$	185.18	188.61	191.57	188.71
p_D	146.20	144.09	140.83	141.59

Table 2.4: Serious vascular events prevention data: comparison of posterior mean and 95% credible interval under 7 different models, specifically marginal log odds ratio mLOR_{ij} comparing the i^{th} and j^{th} treatments, absolute risk of events for the t^{th} treatment (p_t), standard deviation of log-odds for t^{th} treatment (δ_t), and correlation between treatment-specific (treatments i and j) log-odds (ρ_{ij}). Regimen labels: (1) placebo, (2) aspirin plus thienopyridines, (3) aspirin, (4) aspirin plus dipyridamole, (5) thienopyridines.

	Point Estimate (95% Credible Interval)						
Parameter	AB-IW	AB-RW	AB-RIW	AB-EQ	AB-RW-EV	AB-RIW-EV	AB-EQ-EV
mLOR_{12}	0.82 (0.03, 1.51)	0.66 (-0.90, 1.49)	0.24 (-1.52, 1.35)	0.14 (-1.32, 0.75)	0.80 (0.28, 1.34)	0.39 (0.16, 0.89)	0.35 (0.16, 0.56)
mLOR_{13}	0.28 (0.00, 0.57)	0.27 (0.06, 0.50)	0.22 (0.08, 0.38)	0.22 (0.08, 0.38)	0.23 (0.04, 0.46)	0.17 (0.07, 0.28)	0.17 (0.07, 0.28)
mLOR_{14}	0.48 (0.06, 0.90)	0.48 (0.16, 0.77)	0.44 (0.26, 0.61)	0.44 (0.27, 0.60)	0.45 (0.15, 0.76)	0.41 (0.27, 0.55)	0.41 (0.28, 0.54)
mLOR_{15}	0.40 (-0.06, 0.85)	0.38 (-0.13, 0.81)	0.30 (0.06, 0.57)	0.30 (0.09, 0.53)	0.36 (0.06, 0.69)	0.25 (0.10, 0.41)	0.25 (0.11, 0.40)
mLOR_{23}	-0.54 (-1.23, 0.23)	-0.39 (-1.24, 1.18)	-0.02 (-1.13, 1.74)	0.08 (-0.50, 1.53)	-0.57 (-1.10, -0.07)	-0.21 (-0.70, -0.01)	-0.18 (-0.36, -0.01)
mLOR_{24}	-0.34 (-1.09, 0.49)	-0.18 (-1.05, 1.39)	0.20 (-0.91, 1.96)	0.30 (-0.31, 1.75)	-0.35 (-0.92, 0.20)	0.02 (-0.48, 0.27)	0.06 (-0.17, 0.26)
mLOR_{25}	-0.42 (-1.20, 0.40)	-0.28 (-1.26, 1.30)	0.06 (-1.09, 1.80)	0.15 (-0.42, 1.57)	-0.43 (-1.03, 0.09)	-0.13 (-0.63, 0.07)	-0.10 (-0.28, 0.06)
mLOR_{34}	0.21 (-0.20, 0.60)	0.21 (-0.11, 0.51)	0.22 (0.03, 0.38)	0.22 (0.05, 0.38)	0.22 (-0.05, 0.50)	0.24 (0.10, 0.37)	0.24 (0.10, 0.36)
mLOR_{35}	0.13 (-0.32, 0.56)	0.11 (-0.40, 0.55)	0.08 (-0.13, 0.31)	0.07 (-0.09, 0.25)	0.13 (-0.15, 0.43)	0.08 (-0.04, 0.21)	0.08 (-0.03, 0.19)
mLOR_{45}	-0.08 (-0.62, 0.46)	-0.10 (-0.64, 0.40)	-0.14 (-0.40, 0.15)	-0.15 (-0.37, 0.10)	-0.09 (-0.45, 0.28)	-0.16 (-0.33, 0.03)	-0.16 (-0.32, 0.02)
p_1	0.21 (0.17, 0.25)	0.21 (0.18, 0.24)	0.20 (0.17, 0.23)	0.20 (0.17, 0.23)	0.20 (0.17, 0.24)	0.19 (0.16, 0.22)	0.19 (0.16, 0.22)
p_2	0.11 (0.06, 0.20)	0.13 (0.06, 0.39)	0.18 (0.06, 0.53)	0.19 (0.11, 0.48)	0.10 (0.06, 0.15)	0.14 (0.09, 0.17)	0.14 (0.12, 0.18)
p_3	0.17 (0.14, 0.20)	0.17 (0.14, 0.20)	0.17 (0.14, 0.20)	0.17 (0.14, 0.20)	0.17 (0.14, 0.19)	0.17 (0.14, 0.19)	0.17 (0.14, 0.19)
p_4	0.14 (0.10, 0.19)	0.14 (0.11, 0.18)	0.14 (0.11, 0.17)	0.14 (0.11, 0.17)	0.14 (0.11, 0.17)	0.14 (0.11, 0.16)	0.14 (0.11, 0.16)
p_5	0.15 (0.11, 0.21)	0.15 (0.11, 0.23)	0.16 (0.12, 0.20)	0.16 (0.12, 0.19)	0.15 (0.12, 0.18)	0.15 (0.13, 0.18)	0.16 (0.13, 0.18)
δ	.	.	.	.	0.37 (0.27, 0.49)	0.42 (0.30, 0.59)	0.42 (0.30, 0.59)
δ_1	0.42 (0.28, 0.65)	0.28 (0.13, 0.49)	0.36 (0.19, 0.61)	0.36 (0.19, 0.62)	.	.	.
δ_2	0.55 (0.30, 1.05)	0.86 (0.14, 3.63)	0.93 (0.19, 3.67)	0.82 (0.23, 2.73)	.	.	.
δ_3	0.52 (0.36, 0.75)	0.46 (0.30, 0.68)	0.51 (0.34, 0.77)	0.52 (0.35, 0.77)	.	.	.
δ_4	0.48 (0.29, 0.79)	0.34 (0.12, 0.74)	0.40 (0.18, 0.73)	0.41 (0.19, 0.73)	.	.	.
δ_5	0.57 (0.35, 0.94)	0.60 (0.29, 1.27)	0.55 (0.31, 1.00)	0.54 (0.32, 0.92)	.	.	.
ρ	.	.	.	0.97 (0.85, 1.00)	.	.	0.98 (0.90, 1.00)
ρ_{12}	0.05 (-0.62, 0.66)	0.07 (-0.71, 0.79)	0.73 (-0.95, 1.00)	.	0.10 (-0.68, 0.80)	0.90 (-0.24, 1.00)	.
ρ_{13}	0.43 (-0.12, 0.79)	0.68 (0.11, 0.96)	0.96 (0.81, 1.00)	.	0.70 (0.17, 0.96)	0.98 (0.88, 1.00)	.
ρ_{14}	0.25 (-0.37, 0.72)	0.42 (-0.33, 0.91)	0.95 (0.68, 1.00)	.	0.48 (-0.28, 0.93)	0.97 (0.83, 1.00)	.
ρ_{15}	0.23 (-0.43, 0.73)	0.33 (-0.43, 0.87)	0.92 (0.54, 1.00)	.	0.40 (-0.31, 0.89)	0.96 (0.78, 1.00)	.
ρ_{23}	0.07 (-0.67, 0.73)	0.10 (-0.68, 0.80)	0.73 (-0.94, 1.00)	.	0.14 (-0.63, 0.81)	0.90 (-0.23, 1.00)	.
ρ_{24}	0.04 (-0.65, 0.68)	0.04 (-0.73, 0.77)	0.72 (-0.94, 1.00)	.	0.08 (-0.70, 0.79)	0.90 (-0.23, 1.00)	.
ρ_{25}	0.07 (-0.67, 0.76)	0.07 (-0.70, 0.80)	0.71 (-0.90, 1.00)	.	0.09 (-0.69, 0.81)	0.90 (-0.22, 1.00)	.
ρ_{34}	0.35 (-0.27, 0.78)	0.50 (-0.18, 0.92)	0.95 (0.69, 1.00)	.	0.56 (-0.10, 0.94)	0.97 (0.83, 1.00)	.
ρ_{35}	0.32 (-0.38, 0.79)	0.39 (-0.34, 0.88)	0.92 (0.56, 1.00)	.	0.48 (-0.15, 0.90)	0.97 (0.79, 1.00)	.
ρ_{45}	0.18 (-0.54, 0.74)	0.21 (-0.59, 0.83)	0.91 (0.44, 1.00)	.	0.28 (-0.49, 0.86)	0.96 (0.73, 1.00)	.
DIC	89.97	88.56	79.99	79.11	87.08	74.91	74.55
$\bar{D}$	48.23	49.77	46.60	46.48	49.71	46.27	46.19
p_D	41.74	38.78	33.40	32.62	37.38	28.64	28.37

Table 2.5: Postpartum haemorrhage prevention data: comparison of posterior mean and 95% credible interval under 7 different models, specifically marginal log odds ratio mLOR_{ij} comparing the i^{th} and j^{th} treatments, absolute risk of events for the t^{th} treatment (p_t), standard deviation of log-odds for t^{th} treatment (δ_t). Regimen labels: (1) other uterotonics, (2) misoprostol 600-800 mcg, (3) misoprostol 400-500 mcg, (4) misoprostol $<$ 400 mcg.

	Point Estimate (95% Credible Interval)						
	AB-IW	AB-RW	AB-RIW	AB-EQ	AB-RW-EV	AB-RIW-EV	AB-EQ-EV
mLOR_{12}	-0.10 (-0.78, 0.57)	-0.21 (-1.38, 0.69)	-0.16 (-0.83, 0.42)	-0.16 (-0.80, 0.38)	-0.13 (-0.81, 0.56)	-0.17 (-0.53, 0.25)	-0.17 (-0.49, 0.21)
mLOR_{13}	-0.21 (-0.73, 0.29)	-0.21 (-0.83, 0.35)	-0.25 (-0.73, 0.20)	-0.26 (-0.76, 0.20)	-0.14 (-0.51, 0.25)	-0.18 (-0.45, 0.11)	-0.18 (-0.45, 0.10)
mLOR_{14}	0.14 (-3.08, 3.39)	-0.70 (-4.44, 2.76)	-0.59 (-4.40, 2.75)	0.73 (-1.27, 3.19)	0.08 (-2.61, 3.26)	0.05 (-2.76, 3.31)	1.13 (-0.74, 3.78)
mLOR_{23}	-0.11 (-0.87, 0.68)	0.00 (-1.01, 1.24)	-0.09 (-0.78, 0.67)	-0.10 (-0.77, 0.64)	-0.01 (-0.75, 0.73)	-0.01 (-0.47, 0.43)	-0.02 (-0.44, 0.39)
mLOR_{24}	0.23 (-3.00, 3.52)	-0.48 (-4.25, 3.15)	-0.44 (-4.24, 2.96)	0.89 (-1.17, 3.42)	0.22 (-2.53, 3.46)	0.21 (-2.60, 3.50)	1.30 (-0.62, 3.98)
mLOR_{34}	0.35 (-2.87, 3.62)	-0.49 (-4.20, 3.02)	-0.34 (-4.18, 3.05)	0.99 (-1.07, 3.47)	0.22 (-2.48, 3.41)	0.23 (-2.58, 3.49)	1.32 (-0.57, 3.96)
p_1	0.05 (0.03, 0.09)	0.05 (0.03, 0.09)	0.06 (0.03, 0.10)	0.06 (0.03, 0.10)	0.05 (0.03, 0.08)	0.05 (0.03, 0.09)	0.05 (0.03, 0.09)
p_2	0.06 (0.03, 0.12)	0.07 (0.03, 0.17)	0.07 (0.03, 0.14)	0.07 (0.03, 0.14)	0.06 (0.03, 0.11)	0.06 (0.03, 0.11)	0.06 (0.03, 0.11)
p_3	0.06 (0.03, 0.12)	0.06 (0.03, 0.12)	0.07 (0.04, 0.14)	0.07 (0.04, 0.14)	0.06 (0.03, 0.09)	0.06 (0.04, 0.11)	0.06 (0.04, 0.11)
p_4	0.09 (0.00, 0.52)	0.18 (0.00, 0.80)	0.19 (0.00, 0.79)	0.04 (0.00, 0.18)	0.09 (0.00, 0.39)	0.10 (0.00, 0.43)	0.03 (0.00, 0.10)
δ	.	.	.	.	1.15 (0.82, 1.61)	1.25 (0.86, 1.81)	1.25 (0.86, 1.82)
δ_1	1.23 (0.82, 1.81)	1.20 (0.80, 1.79)	1.32 (0.86, 2.01)	1.32 (0.86, 2.00)	.	.	.
δ_2	1.17 (0.64, 2.00)	1.33 (0.64, 2.84)	1.32 (0.71, 2.38)	1.33 (0.71, 2.40)	.	.	.
δ_3	1.32 (0.82, 2.08)	1.34 (0.81, 2.18)	1.46 (0.88, 2.34)	1.48 (0.89, 2.37)	.	.	.
δ_4	1.14 (0.38, 3.19)	2.48 (0.16, 4.86)	2.50 (0.13, 4.88)	2.48 (0.19, 4.85)	.	.	.
DIC	68.83	70.62	68.03	67.86	69.25	65.09	64.68
$\bar{D}$	39.42	40.10	40.66	40.30	39.97	39.82	39.70
p_D	29.41	30.52	27.37	27.55	29.28	25.28	24.98

Chapter 3

A Variance Shrinkage Method Improves Arm-Based Bayesian Network Meta-Analysis

3.1 Introduction

As mentioned in Section 1.2.1, lack of information is a big problem in AB-NMA. In this chapter, we introduce a variance shrinkage method to solve this problem. Specifically, we assume different treatment-specific variances share a common prior with unknown hyper-parameters. This assumption is weaker than the homogeneous-variance assumption and improves estimation by shrinking the variances in a data-dependent way.

The rest of this chapter is organized as follows. Section 3.2 describes a motivating example of an NMA of organized inpatient care for stroke. Section 3.3 gives a brief review of the AB model for analyzing NMA datasets with dichotomous outcomes and introduces the variance shrinkage method. Section 3.4 presents results from applying the variance shrinkage method to the motivating example, followed by extensive simulation studies in Section 3.5, comparing the performance of different priors on variances. Section 3.6 presents our findings with a brief discussion.

3.2 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 [61] and the third leading cause of disability [62]. 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 [63]. 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 [64] 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 3.1 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 3.1. We reanalyzed the stoke data using the AB-NMA approach specified by Zhang et al. [7] Specifically, we used the exchangeable correlation structure with a uniform prior $U(-\frac{1}{T-1}, 1)$ on the correlation coefficients [65] and the heterogeneous variance assumption with separate uniform priors $U(0, 5)$ on the standard deviations (henceforth referred to as the

UV approach). Figures 3.2A and 3.2B present the forest plots of the standard deviations and absolute risks of treatments 1–5. 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.

3.3 Methods

3.3.1 Arm-based Bayesian network meta-analysis

This subsection gives a brief introduction to AB-NMA [7]. 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:

$$\begin{aligned} &\text{Level I: } r_{kt} \sim \text{Binomial}(n_{kt}, p_{kt}),\ t \in A_k,\ k = 1, \ldots, K; \\ &\text{Level II: } \text{logit}(p_{kt}) = \theta_{kt}; \quad (\theta_{k1}, \ldots, \theta_{kT})' \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned} \tag{3.1}$$

where p_{kt} is the probability of an event (i.e., absolute risk) 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.

3.3.2 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 $\boldsymbol{\Sigma}$, we use the separation strategy proposed by Barnard et al. [17] Instead of treating the covariance matrix as a whole, this method first decomposes it into separate parts as $\boldsymbol{\Sigma} = \boldsymbol{\Delta}\mathbf{P}\boldsymbol{\Delta}$ 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 $\boldsymbol{\Delta}$. Here, we will simply use the exchangeable correlation structure to set a prior for the correlation matrix $\mathbf{P}$ [47]; 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.

3.3.3 Variance shrinkage method

As mentioned in Section 3.2, 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 hyper-parameters, and we use the data to estimate the hyper-parameters.

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 hyper-parameter a, which has a uniform prior $U(\epsilon_l, \epsilon_u)$. The HC prior is commonly used for standard deviations [19] 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 3.3). By using the HHC prior, the data-driven posterior distribution of the hyper-parameter a determines how informative the prior on δ_t should be: if it is less informative (e.g., $a = 5$ as in Figure 3.3), it shrinks δ_t less; if it is more informative (e.g., $a = 0.5$ as in Figure 3.3), it shrinks δ_t more.

3.3.4 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:

$$L(\boldsymbol{\theta}_k|D_k) = \prod_{t\in A_k} [\text{logit}^{-1}(\theta_{kt})]^{r_{kt}}[1-\text{logit}^{-1}(\theta_{kt})]^{n_{kt}-r_{kt}}. \tag{3.2}$$

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 $p(\boldsymbol{\theta}_k|\boldsymbol{\mu},\boldsymbol{\Sigma}) = p(\boldsymbol{\theta}_k|\boldsymbol{\mu},\boldsymbol{\Delta},\rho)$, the joint posterior distribution is:

$$\begin{aligned}
&\pi(\boldsymbol{\mu},\boldsymbol{\Delta},a,\rho,\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_K|D) \propto \\
&\prod_{k=1}^{K}\left\{\prod_{t\in A_k}[\text{logit}^{-1}(\theta_{kt})]^{r_{kt}}[1-\text{logit}^{-1}(\theta_{kt})]^{n_{kt}-r_{kt}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\boldsymbol{\theta}_k-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_k-\boldsymbol{\mu})}\right\}\times \\
&\prod_{t=1}^{T}\pi(\mu_t)\pi(\delta_t|a)\pi(a)\pi(\rho),
\end{aligned} \tag{3.3}$$

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 (3.1), p_t can be approximated by [44]

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

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.

3.4 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 $\mathbf{\Sigma}$ 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 $U(-\frac{1}{T-1}, 1)$ 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 diffrence" between models 2 and 3. 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. [66] 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:

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

3.4.1 Model comparison

We evaluated the models using the deviance information criterion (DIC) by Spiegelhalter et al. [51] and the widely applicable information criteria (WAIC) [67]. 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. [68] Specifically, DIC is calculated as:

$$\text{DIC} = \overline{D} + p_D, \tag{3.4}$$

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

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

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

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

The other criterion, WAIC, is asymptotically equivalent to leave-one-out cross-validation [67]. 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):

$$\begin{aligned} \text{WAIC} &= -2\overline{\text{lppd}} + 2\overline{p_W}; \\ \text{lppd} &= \sum_{k=1}^{K} \sum_{t \in A_k} \log \int p(r_{kt}, n_{kt} | \boldsymbol{\psi}) \pi(\boldsymbol{\psi} | D) d\boldsymbol{\psi}; \\ p_W &= \sum_{k=1}^{K} \sum_{t \in A_k} \text{Var}(\log(p(r_{kt}, n_{kt} | \boldsymbol{\psi}))), \end{aligned} \tag{3.5}$$

where $\pi(\boldsymbol{\psi}|D)$ is the joint posterior distribution in Equation (3.3), 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

$$\begin{aligned} \overline{\text{lppd}} &= \sum_{k=1}^{K} \sum_{t \in A_k} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(r_{kt}, n_{kt} | \boldsymbol{\psi}^s) \right); \\ \overline{p_W} &= \sum_{k=1}^{K} \sum_{t \in A_k} \frac{1}{S-1} \sum_{s=1}^{S} \left(\log(p(r_{kt}, n_{kt} | \boldsymbol{\psi}^s)) - \overline{\log(p(r_{kt}, n_{kt} | \boldsymbol{\psi}^s))} \right)^2. \end{aligned} \tag{3.6}$$

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

3.4.2 Results

Appendix A.1 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 3.1 presents results for the marginal log odds ratios mLOR_{ij}, the absolute risk (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 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 3.2A 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 absolute risks, as shown in Figure 3.2B. 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 absolute risk; 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.

3.5 Simulation studies

3.5.1 Simulation settings

We conducted comprehensive simulation studies to compare the four priors defined and used in Section 3.4. 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 (3.1) 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 3.4 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 3.4. 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 3.4.

3.5.2 Simulation results

Table 3.2 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 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, MAR2). We evaluated the log odds ratio comparing treatments i and j (mLOR_{ij} and cLOR_{ij}), the absolute risk of treatment t (p_t), the standard deviation for treatment t (δ_t), and the correlation between treatments i and j (ρ_{ij}). Due to space limits, instead of presenting the results for each treatment comparison, for each of $\text{Bias}_{\bar{\mu}}$, $\text{Bias}_{\tilde{\mu}}$ and $\text{MSE}_{\tilde{\mu}}$, we calculated the sum of the absolute value over all pairs of comparisons. For example, the entry in Table 3.2 with $\text{Bias}_{\bar{\mu}}$ as the column and cLOR_{ij} as the row was calculated as $\sum_{i\neq j} |\text{Bias}_{\bar{\mu}}(\text{cLOR}_{ij})|$. To summarize the CPs, the corresponding value in Table 3.2 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 3.3 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 4 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 absolute risks 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 absolute

risks 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.

3.6 Summary and discussion

This chapter 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 4 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 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. [69] and Ghosh et al. [70], we considered the Student-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 3.4; 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 3.4 summarizes the pros and cons of these methods. The UV prior leads to biased estimates and the credible intervals did not have nominal CP when B_t was small ($\leq$4). The IW prior could not estimate correlations and standard deviations accurately, which resulted in biased log odds ratios and absolute risks 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 coverage. The HHC prior generally had the best performance among the 4 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 chapter focused on shrinking standard deviations in the AB-NMA with binary outcomes; many extensions are possible. First, the HHC is just one choice of prior for inducing shrinkage. Other priors, such as the hierarchical inverse-gamma (HIG) prior, denoted by HIG(ϵ_l, ϵ_u), could be considered. The HIG prior's density is $p(\delta_t^2|\beta) \propto$ IG(α, β) 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 IG$(1, 0.1)$, and a prior that may overestimate a variance with true value close to zero, such as IG$(1, 1)$; see Figure 3.3. We compared HIG$(0, 1)$ with other four methods (see Appendix A.2) 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."

Second, while we chose $\epsilon_l = 0$ and $\epsilon_u = 5$ in the uniform prior for the HHC's hyper-parameter 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 hyper-parameters, 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 multi-arm (≥ 2) studies versus single-arm studies. Therefore, methods for combining single-arm and multiple-arm studies should be further examined.

Figure 3.1: 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).

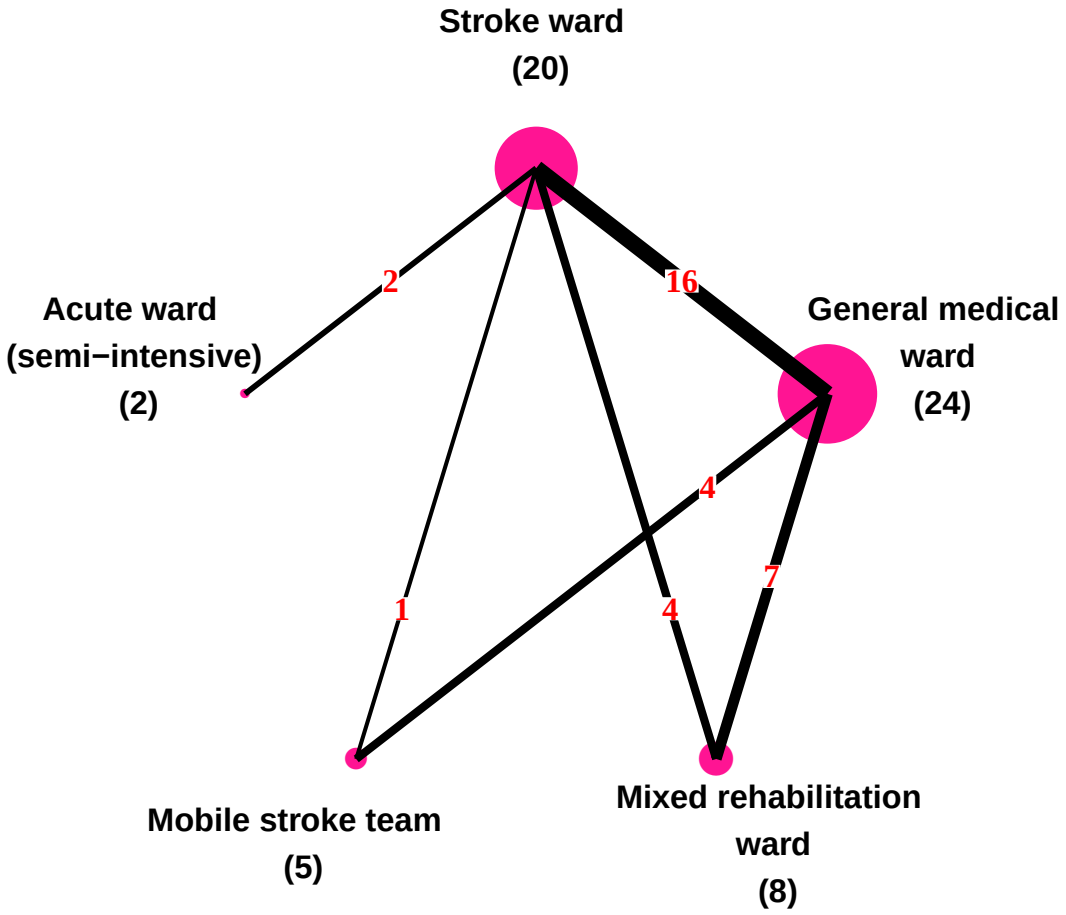
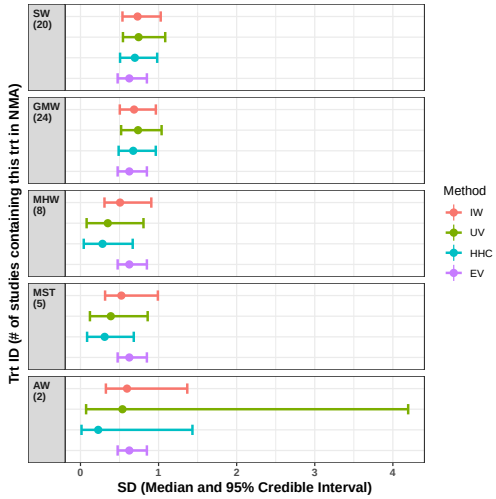


Figure 3.2: Results for case study of organized inpatient care for stroke: Forest plot of standard deviations δ_t and absolute risk p_t (posterior median with 95% credible interval). Different colors indicate different priors. The y-axis represents the treatment label, with B_t in parentheses. Treatment labels: 1) stroke ward, 2) general medical ward, 3) mixed rehabilitation ward, 4) mobile stroke team, and 5) acute (semi-intensive) ward.

(A) standard deviation δ_t



(B) absolute risk p_t

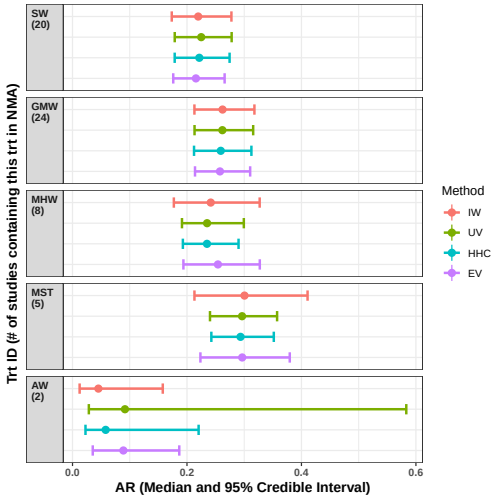


Figure 3.3: Densities of different priors on standard deviation δ_t. For better visualization, the horizontal axis is limited to $[0, 5]$.

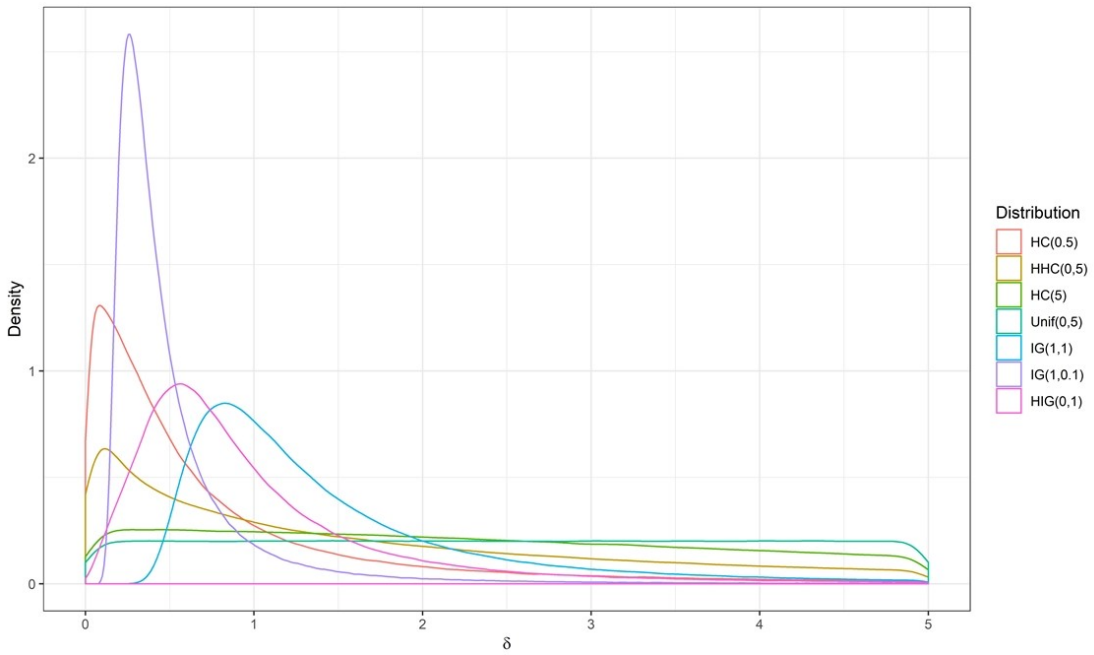


Table 3.1: Organized inpatient care for stroke data: Comparing posterior median and 95% credible intervals under 4 models (IW, UV, HHC and EV); mLOR_{ij} compares the i^{th} and j^{th} treatment, absolute risk of events for the t^{th} treatment (p_t), standard deviation of the t^{th} treatment (δ_t), and SUCRA of the t^{th} treatment (SUCRA_t). Treatment labels: 1) stroke ward, 2) general medical ward, 3) mixed rehabilitation ward, 4) mobile stroke team, and 5) acute (semi-intensive) ward.

	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
$\bar{D}$	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

Figure 3.4: Missing data structures for simulation study. (a) MCAR1 and MAR1; (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.

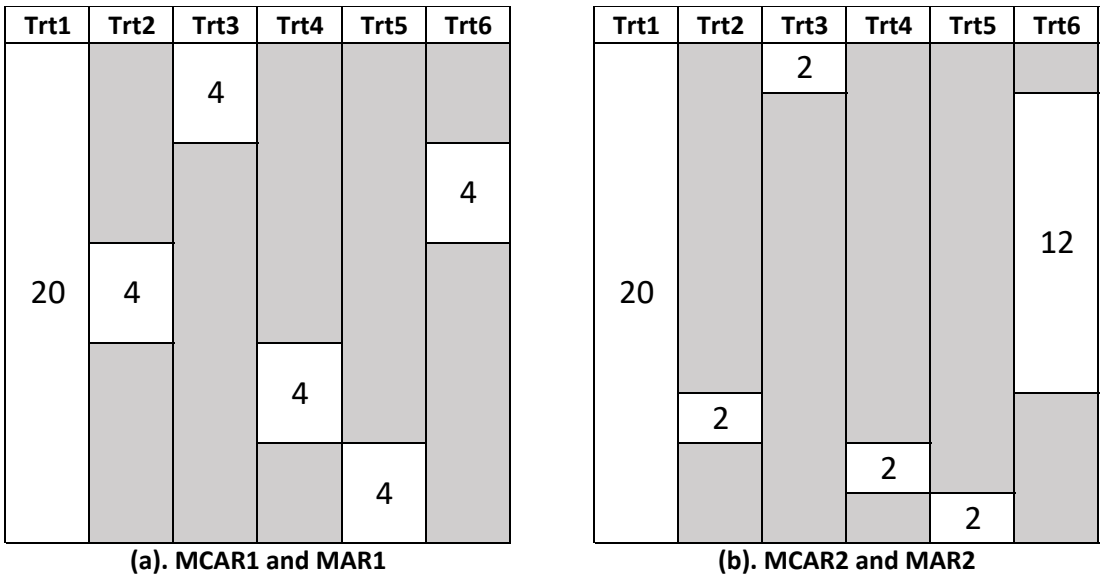


Table 3.2: Simulation results for data generated under scenario I (heterogeneous variance) with 4 different missingness settings (MCAR1, MCAR2, MAR1, MAR2), specifically bias of the posterior mean ($\text{Bias}_{\bar{\mu}}$), bias of the posterior median ($\text{Bias}_{\tilde{\mu}}$), mean squared error of the posterior median ($\text{MSE}_{\tilde{\mu}}$), and coverage probability (CP) of the 95% credible intervals for 4 different priors. 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 $\sum_{i\neq j}(0.95 - \text{CP}(\text{cLOR}_{ij}))_+$.

		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

Table 3.3: Simulation results comparing data generated under scenario II (homogeneous variance) with 4 different missingness settings (MCAR1, MCAR2, MAR1, MAR2). 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 4 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				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

Table 3.4: Pros and cons of the four different models.

Model	Computing burden	Performance					
		MCAR			MAR		
		All B_ts are large	Small B_t exists		All B_ts are large	Small B_t exists	
			δ_ts are similar	δ_ts differ		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

Chapter 4

Bridging randomized controlled trials and single-arm studies using commensurate priors in arm-based network meta-analysis

4.1 Introduction

As discussed in Section 1.2.2, an extrapolation strategy in meta-analysis and NMA is a natural solution to the lack of information. Several statistical methods have been developed for "information borrowing" in the field of Bayesian analysis, for instance, power priors [71–73], hierarchical commensurate priors [74–76], and Bayesian model averaging [77–79]. Motivated by these methods, we propose commensurate priors to adaptively incorporate *variance* information from single-arm studies into an AB-NMA. Although an AB-NMA naturally incorporates single-arm studies [16], it does not explicitly account for the possibly lower quality of single-arm studies. Our new method, by contrast, has the advantage of downweighting single-arm studies when they appear to be inconsistent with the NMA's two- or multi-arm studies.

The rest of this chapter is organized as follows. Section 4.2 describes a motivating example, an NMA comparing safety of different immune checkpoint inhibitors for treating

cancer. Section 4.3 introduces commensurate priors to combine RCTs and single-arm studies in an AB-NMA. Section 4.4 presents results from applying our method to the motivating example, followed by simulation studies in Section 4.5, comparing the performance of different commensurate priors. Section 4.6 summarizes our findings with a brief discussion.

4.2 Motivating example: safety of immune checkpoint inhibitors in cancer

Immune checkpoint inhibitor (ICI) has recently emerged as a breakthrough in treating more than 14 different cancers, including melanoma, Hodgkin lymphoma, non-small cell lung cancer, and others [80]. Immune checkpoints (also known as receptors on T-cells) can prevent autoimmunity of T-cells during encounter with tumor cells or tumor-associated antigen-presenting cells by sequestering CD80/CD86 ligands that would otherwise signal through CD28 (CTAL-4: cytotoxic T lymphocyte associated antigen 4) or by inducing T-cell exhaustion (PD-1/PD-L1: programmed cell death 1/programmed cell death ligand 1) [81]. As monoclonal antibodies, ICIs can target these immune checkpoints and remove inhibition of T-cell function. However, they may also promote T-cell activity against host tissues, facilitating autoimmune activity against any organ in the body, which raises a concern about tolerability. To investigate the safety of ICIs, Xu et al. [82] conducted a systematic review and NMA mainly on five ICIs, including ipilimumab, tremelimumab (both against CTAL-4), nivolumab, pembrolizumab (both against PD-1), and atezolizumab (against PD-L1). Only 3 out of 31 RCTs had an atezolizumab arm, making parameters related to atezolizumab (variance, absolute risks, and relative risks) difficult to estimate. Fortunately, Xu et al. [82] selected not only the 31 RCTs but also found 36 single-arm studies as a validation group. We collected all available data from these RCTs and single-arm studies for our analysis.

Table 4.1 shows the modified dataset of 27 RCTs and 28 single-arm studies comparing eight treatments: 1) nivolumab 3mg/kg every 2 weeks (NIV); 2) ipilimumab 3mg/kg every 3 weeks (IPI low); 3) ipilimumab 10mg/kg every 3 weeks (IPI high); 4) pembrolizumab (PEM); 5) atezolizumab 1200mg every 3 weeks (ATE); 6) one ICI drug plus investigator's choice chemotherapy (ICI+ICC); 7) two ICI drugs together (2ICIs);

and 8) investigator's choice chemotherapy (ICC). The outcome is safety, specifically occurrence of any treatment-related grade 3-5 adverse events. We excluded 9 single-arm studies because they did not provide information about treatment-related adverse events. Because only one RCT investigated tremelimumab, which did not show a statistically significant survival advantage over standard chemotherapy in first-line treatment of patients with metastatic melanoma [83] and also did not have any available information from single-arm studies, our analysis excluded tremelimumab. We also excluded 2 RCTs [84, 85] and transformed one RCT [86] into a single-arm study (by dropping one treatment arm) because the doses of ICIs investigated in these trials (e.g., nivolumab 0.3mg/kg every 3 weeks, nivolumab 2mg/kg every 3 weeks, and nivolumab 10mg/kg every 3 weeks in Motzer at al.'s [84] trial did not match doses in the other RCTs. Figure 4.1 intuitively shows the need to borrow information from single-arm studies to potentially improve estimation. For example, only 4 and 3 RCTs contain ipilimumab (high dose) and atezolizumab respectively, which causes difficulty in estimating these variances of these two treatment-specific log-odds; however, with the additional information from single-arm studies (3 for each of these two ICIs), we may be able to overcome this problem.

4.3 Statistical methods

4.3.1 Notation

Assume an NMA has K clinical trials comparing a total of T treatments. Let $\mathcal{A}_k$ $(k = 1, \ldots, K)$ be the subset of treatments in the k^{th} study. For most clinical trials, the number of treatments in $\mathcal{A}_k$ (denoted by $|\mathcal{A}_k|$) is 2 or 3. Let $\mathcal{D}_k$ be the data observed in the k^{th} RCT. For NMAs with a dichotomous outcome, $\mathcal{D}_k = \{(r_{kt}, n_{kt}), t \in \mathcal{A}_k\}$, where r_{kt} and n_{kt} are the numbers of events and participants respectively for the t^{th} treatment in the k^{th} RCT. Assume that the NMA also includes J high-quality single-arm studies. Let $\mathcal{D}^{\text{s}}_j$ $(j = 1, \ldots, J)$ be the data collected in the j^{th} single-arm study; $\mathcal{D}^{\text{s}}_j = \{(r^{\text{s}}_{jt}, n^{\text{s}}_{jt}), t \in \mathcal{A}^{\text{s}}_j\}$, where $\mathcal{A}^{\text{s}}_j$ includes only one treatment ($|\mathcal{A}^{\text{s}}_j| = 1$) and r^{s}_{jt} and n^{s}_{jt} are the numbers of events and participants for the t^{th} treatment in the j^{th} single-arm study. We further define B_t to be the number of clinical trials containing the t^{th} treatment, and B^{s}_t to be the number of single-arm studies containing the t^{th}

treatment. For instance, as shown in Figure 4.1, 9 RCTs and 11 single-arm studies contain nivolumab, so $B_1 = 9$ and $B_1^{\mathrm{s}} = 11$.

4.3.2 Arm-based network meta-analysis and model for single-arm studies

This subsection briefly introduces the AB-NMA [7, 11] and the model for single-arm studies. This chapter focuses on the case of binary outcomes. The underlying model is:

$$\begin{aligned} & r_{kt} \sim \text{Binomial}(n_{kt}, p_{kt}),\ t \in \mathcal{A}_k,\ k = 1, \ldots, K; \\ & \text{logit}(p_{kt}) = \theta_{kt}; \\ & (\theta_{k1}, \ldots, \theta_{kT})' \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned} \tag{4.1}$$

where p_{kt} is the probability of an event (i.e., absolute risk) for the t^{th} treatment in the k^{th} study and the latent log odds $\boldsymbol{\theta}_k = (\theta_{k1}, \ldots, \theta_{kT})'$ are assumed to follow the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Here, $\boldsymbol{x}'$ denotes the transpose of the vector $\boldsymbol{x}$. The vector of latent variables $\boldsymbol{\theta}_k$ models all T treatments, even though only $|\mathcal{A}_k|$ treatments t are actually observed in study k. Moreover, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)'$ contains the overall logit event probability for each treatment. If we denote the between-study standard deviation of 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. We further define $\boldsymbol{\mu}^{\mathrm{n}} = (\mu_1, \ldots, \mu_T)'$, and $\boldsymbol{\delta}^{\mathrm{n}} = (\delta_1, \ldots, \delta_T)'$. We call this original method *no borrowing* (NB) because it does not incorporate any information from single-arm studies.

Similarly, the model for single-arm studies is:

$$\begin{aligned} & r_{jt}^{\mathrm{s}} \sim \text{Binomial}(n_{jt}^{\mathrm{s}}, p_{jt}^{\mathrm{s}}),\ t \in \mathcal{A}_j^{\mathrm{s}},\ j = 1, \ldots, J; \\ & \text{logit}(p_{jt}^{\mathrm{s}}) = \theta_{jt}^{\mathrm{s}}; \\ & \theta_{jt}^{\mathrm{s}} \sim N(\mu_t^{\mathrm{s}}, (\delta_t^{\mathrm{s}})^2), \end{aligned} \tag{4.2}$$

where p_{jt}^{s} is the probability of an event for the j^{th} single-arm study, μ_t^{s} represents the overall fixed effect of treatment t from single-arm studies, and δ_t^{s} is the standard deviation of the t^{th} treatment for single-arm studies. Unlike Equation (4.1) containing some latent variables that correspond to unobserved treatment arms, the variables in

Equation (4.2) all correspond to observed treatment arms. We further define $\boldsymbol{\mu}^{\mathrm{s}} = (\mu_1^{\mathrm{s}}, \ldots, \mu_T^{\mathrm{s}})'$, and $\boldsymbol{\delta}^{\mathrm{s}} = (\delta_1^{\mathrm{s}}, \ldots, \delta_T^{\mathrm{s}})'$.

4.3.3 Connecting NMA and single-arm studies

With models in hand for an NMA and single-arm studies, we consider several methods to adaptively integrate information from the single-arm studies into the NMA.

Existing methods: full borrowing

The AB-NMA model in Equation (4.1) could naturally incorporate information from single-arm studies about means and variances by assuming $\mu_t = \mu_t^{\mathrm{s}}$ and $\delta_t = \delta_t^{\mathrm{s}}$, $t = 1, \ldots, T$. We call this method *fully borrowing on means and variances* (FBMV). However, this assumption may be too strong and unrealistic. Instead, we can take a step back and only borrow information about variances from single-arm studies by assuming $\delta_t = \delta_t^{\mathrm{s}}$ while $\mu_t \neq \mu_t^{\mathrm{s}}$. We call this method *fully borrowing on variances* (FBV). We could also borrow mean information only by assuming $\mu_t = \mu_t^{\mathrm{s}}$ while $\delta_t \neq \delta_t^{\mathrm{s}}$, but this chapter will not discuss this *fully borrowing on mean* (FBM) method in detail as it might be less practical.

Commensurate prior on mean

Although a fully borrowing approach could naturally integrate single-arm studies, it may also cause large biases if the reliability of single-arm studies may be doubtful. Instead, a commensurate prior on the means, introduced by Hobbs et al. [75], is a simple way to offer flexibility in borrowing from and downweighting single-arm studies:

$$\mu_t \sim N(\mu_t^{\mathrm{s}}, \eta^{-1}), \tag{4.3}$$

where μ_t has a normal prior with mean μ_t^{s} and precision η. The precision η characterizes how commensurate the two sources of information (μ_t and μ_t^{s}) are with each other. Hobbs et al. [75] proposed a "spike-and-slab" prior for η, but the estimation of η is still difficult with this prior. Instead, Murray et al. [76] proposed a modified commensurate

prior:

$$\begin{aligned}&\mu_t \sim [N(\mu_t^{\mathrm{s}}, (\tau_t^{\mathrm{m}})^{-1})]^{1-\iota_t^{\mathrm{m}}} [N(\mu_t^{\mathrm{s}}, (R^{\mathrm{m}})^{-1})]^{\iota_t^{\mathrm{m}}},\\ &\iota_t^{\mathrm{m}} \sim \mathrm{Bern}(p^{\mathrm{m}}) \text{ and } \tau_t^{\mathrm{m}} \sim U(s_{\mathrm{l}}^{\mathrm{m}}, s_{\mathrm{u}}^{\mathrm{m}}),\ t = 1, \ldots, T;\end{aligned} \tag{4.4}$$

where $\mathrm{Bern}(p^{\mathrm{m}})$ denotes a Bernoulli distribution with $\Pr(\iota_t^{\mathrm{m}} = 1) = p^{\mathrm{m}}$, and the prior distribution on the precision τ_t^{m} is uniform from $s_{\mathrm{l}}^{\mathrm{m}}$ to $s_{\mathrm{u}}^{\mathrm{m}}$ with $0 \leq s_{\mathrm{l}}^{\mathrm{m}} < s_{\mathrm{u}}^{\mathrm{m}} \ll R^{\mathrm{m}}$ and $0 \leq p_t^{\mathrm{m}} \leq 1$ pre-specified. With this prior, μ_t follows a two-part mixture normal distribution consisting of a highly concentrated component, i.e., $N(\mu_t^{\mathrm{s}}, (R^{\mathrm{m}})^{-1})$, and a relatively diffuse component, i.e., $N(\mu_t^{\mathrm{s}}, (\tau_t^{\mathrm{m}})^{-1})$. This distribution imitates a "spike-and-slab" prior by putting probability p_t^{m} at a point (i.e., the 'spike' part) to encourage borrowing information from single-arm studies (i.e., μ_t^{s}) and the remaining probability $1 - p_t^{\mathrm{m}}$ on a 'slab' of values close to the original information from the NMA. We call this method *commensurate prior on mean* (CPM).

Commensurate prior on variance

Similarly, we propose a commensurate prior on variances to borrow only variance information from single-arm studies. Specifically, we assume:

$$\log(\delta_t/\delta_t^{\mathrm{s}}) = c_t;\ c_t \sim N(0, \eta^{-1}), \tag{4.5}$$

so the log of the standard deviation ratio follows a normal distribution with mean zero and precision η. Like Murray et al. [76], we can modify this prior as follows:

$$\begin{aligned}&\log(\delta_t) \sim [N(\log(\delta_t^{\mathrm{s}}), (\tau_t^{\mathrm{v}})^{-1})]^{1-\iota_t^{\mathrm{v}}} [N(\log(\delta_t^{\mathrm{s}}), (R^{\mathrm{v}})^{-1})]^{\iota_t^{\mathrm{v}}},\\ &\iota_t^{\mathrm{v}} \sim \mathrm{Bern}(p^{\mathrm{v}}) \text{ and } \tau_t^{\mathrm{v}} \sim U(s_{\mathrm{l}}^{\mathrm{v}}, s_{\mathrm{u}}^{\mathrm{v}}),\ t = 1, \ldots, T;\end{aligned} \tag{4.6}$$

where $\mathrm{Bern}(p^{\mathrm{v}})$ is a Bernoulli distribution with $\Pr(\iota_t^{\mathrm{m}} = 1) = p^{\mathrm{v}}$ and the prior distribution on the precision τ_t^{v} is uniform from $s_{\mathrm{l}}^{\mathrm{v}}$ to $s_{\mathrm{u}}^{\mathrm{v}}$ with $0 \leq s_{\mathrm{l}}^{\mathrm{v}} < s_{\mathrm{u}}^{\mathrm{v}} \ll R^{\mathrm{v}}$ and $0 \leq p_t^{\mathrm{v}} \leq 1$ pre-specified. We call this method *commensurate prior on variance* (CPV). Unlike the FBV method, this prior borrows variance information from single-arm studies adaptively. More specifically, this model encourages borrowing variance information (i.e., δ_t^{s}) from single-arm studies if p^{v} approaches 1, while it tends to ignore single-arm studies if p^{v} approaches 0.

Double commensurate prior

We can borrow both mean and variance information adaptively by applying both the CPV and CPM methods in Equations (4.4) and (4.6). We call this method *commensurate prior on mean and variance* (CPMV) or *double commensurate prior* (DCP); it is an adaptively borrowing version of the FBMV method.

Summary of prior specifications and models

Table 4.2 lists model names, assumptions, and prior specifications in detail. For all these models, we specify a prior on the covariance matrix $\mathbf{\Sigma}$ using the separation strategy proposed by Barnard et al. [17]. Specifically, we first decompose $\mathbf{\Sigma}$ into separate parts as $\mathbf{\Sigma} = \mathbf{\Delta P \Delta}$ and then set priors independently on the correlation matrix $\mathbf{P}$ and the standard deviations δ_t $(t = 1, \ldots, T)$, which are the diagonal elements of $\mathbf{\Delta}$. Here, we focus on the exchangeable correlation prior for the correlation matrix $\mathbf{P}$ [47]: we assume all correlation coefficients ρ_{ij} are equal, i.e., $\rho_{ij} = \rho$ for any $i \neq j$, and assign a uniform prior $U(-\frac{1}{T-1}, 1)$ to ρ so $\mathbf{P}$ is positive-definite. For models in which mean or variance information is not shared between the RCTs and single-arm studies in specific models, we also assign a vague $N(0, 100^2)$ prior to μ_t and μ_t^{s}, and a uniform prior $U(0, 5)$ to δ_t and δ_t^{s}. On the other hand, if information is shared fully or adaptively between the RCTs and single-arm studies, we assume $\mu_t = \mu_t^{\mathrm{s}}$ and $\delta_t = \delta_t^{\mathrm{s}}$ for fully borrowing, or for adaptively borrowing, following Equations (4.4) and (4.6) with pre-specified values $(0.5, 2500, 0, 2)$ for $(p^{\mathrm{m}}, R^{\mathrm{m}}, s_{\mathrm{l}}^{\mathrm{m}}, s_{\mathrm{u}}^{\mathrm{m}})$ and $(p^{\mathrm{v}}, R^{\mathrm{v}}, s_{\mathrm{l}}^{\mathrm{v}}, s_{\mathrm{u}}^{\mathrm{v}})$.

4.3.4 Likelihood and posterior estimation

The likelihood function for $\boldsymbol{\theta}_k$ based on data $\mathcal{D}_k$ from the k^{th} RCT can be written as:

$$L(\boldsymbol{\theta}_k | \mathcal{D}_k) = \prod_{t \in \mathcal{A}_k} [\mathrm{logit}^{-1}(\theta_{kt})]^{r_{kt}} [1 - \mathrm{logit}^{-1}(\theta_{kt})]^{n_{kt} - r_{kt}}. \tag{4.7}$$

Similarly, the likelihood function for θ_{jt}^{s} based on data $\mathcal{D}_j^{\mathrm{s}}$ from the j^{th} single-arm study is:

$$L(\theta_{jt}^{\mathrm{s}} | \mathcal{D}_j^{\mathrm{s}}) = \prod_{t \in \mathcal{A}_j^{\mathrm{s}}} [\mathrm{logit}^{-1}(\theta_{jt}^{\mathrm{s}})]^{r_{jt}^{\mathrm{s}}} [1 - \mathrm{logit}^{-1}(\theta_{jt}^{\mathrm{s}})]^{n_{jt}^{\mathrm{s}} - r_{jt}^{\mathrm{s}}}. \tag{4.8}$$

Since $|\mathcal{A}_j^{\mathrm{s}}| = 1$, we can simply denote θ_{jt}^{s} as θ_j^{s}. Without loss of generality, we focus on illustrating the joint posterior distribution under the DCP model. Noting that $\boldsymbol{\Sigma}$ depends on $\boldsymbol{\Delta}$, and ρ, the joint posterior distribution can be written as:

$$
\begin{aligned}
&\pi(\boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\mu}^{\mathrm{s}}, \boldsymbol{\delta}^{\mathrm{s}}, \rho, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \boldsymbol{\theta}^{\mathrm{s}}, \boldsymbol{\iota}^{\mathrm{m}}, \boldsymbol{\tau}^{\mathrm{m}}, \boldsymbol{\iota}^{\mathrm{v}}, \boldsymbol{\tau}^{\mathrm{v}} | \mathcal{D}_{1:K}, \mathcal{D}_{1:J}^{\mathrm{s}}) \\
&\propto \prod_{k=1}^{K} \left\{ \prod_{t \in \mathcal{A}_k} [\mathrm{logit}^{-1}(\theta_{kt})]^{r_{kt}} [1 - \mathrm{logit}^{-1}(\theta_{kt})]^{n_{kt} - r_{kt}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\theta}_k - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_k - \boldsymbol{\mu})} \right\} \times \\
&\prod_{j=1}^{J} \left\{ \prod_{t \in \mathcal{A}_j^{\mathrm{s}}} [\mathrm{logit}^{-1}(\theta_{jt}^{\mathrm{s}})]^{r_{jt}^{\mathrm{s}}} [1 - \mathrm{logit}^{-1}(\theta_{jt}^{\mathrm{s}})]^{(n_{jt}^{\mathrm{s}} - r_{jt}^{\mathrm{s}})} (\delta_t^{\mathrm{s}})^{-1} e^{-\frac{1}{2}\left(\frac{\theta_{jt}^{\mathrm{s}} - \mu_t^{\mathrm{s}}}{\delta_t^{\mathrm{s}}}\right)^2} \right\} \times \\
&\prod_{t=1}^{T} \left\{ [(\tau_t^{\mathrm{m}})^{\frac{1}{2}} e^{-\frac{1}{2}\tau_t^{\mathrm{m}}(\mu_t - \mu_t^{\mathrm{s}})^2}]^{(1-\iota_t^{\mathrm{m}})} [(R^{\mathrm{m}})^{\frac{1}{2}} e^{-\frac{1}{2}R^{\mathrm{m}}(\mu_t - \mu_t^{\mathrm{s}})^2}]^{\iota_t^{\mathrm{m}}} \right\} \times \\
&\prod_{t=1}^{T} \left\{ \left[(\tau_t^{\mathrm{v}})^{\frac{1}{2}} \delta_t^{-1} e^{-\frac{1}{2}\tau_t^{\mathrm{v}}(\log(\delta_t) - \log(\delta_t^{\mathrm{s}}))^2} \right]^{1-\iota_t^{\mathrm{v}}} \left[(R^{\mathrm{v}})^{\frac{1}{2}} \delta_t^{-1} e^{-\frac{1}{2}R^{\mathrm{v}}(\log(\delta_t) - \log(\delta_t^{\mathrm{s}}))^2} \right]^{\iota_t^{\mathrm{v}}} \right\} \times \\
&\prod_{t=1}^{T} [\pi(\mu_t^{\mathrm{s}}) \pi(\delta_t^{\mathrm{s}}) \pi(\iota_t^{\mathrm{m}}) \pi(\tau_t^{\mathrm{m}}) \pi(\iota_t^{\mathrm{v}}) \pi(\tau_t^{\mathrm{v}})] \times \pi(\rho),
\end{aligned}
\tag{4.9}
$$

where $\mathcal{D}_{1:K} = \{\mathcal{D}_1, \ldots, \mathcal{D}_K\}$, $\mathcal{D}_{1:J}^{\mathrm{s}} = \{\mathcal{D}_1^{\mathrm{s}}, \ldots, \mathcal{D}_J^{\mathrm{s}}\}$, $\boldsymbol{\theta}^{\mathrm{s}} = (\theta_1^{\mathrm{s}}, \ldots, \theta_J^{\mathrm{s}})'$, $\boldsymbol{\iota}^{\mathrm{m}} = (\iota_1^{\mathrm{m}}, \ldots, \iota_T^{\mathrm{m}})'$, $\boldsymbol{\tau}^{\mathrm{m}} = (\tau_1^{\mathrm{m}}, \ldots, \tau_T^{\mathrm{m}})'$, $\boldsymbol{\iota}^{\mathrm{v}} = (\iota_1^{\mathrm{v}}, \ldots, \iota_T^{\mathrm{v}})'$, and $\boldsymbol{\tau}^{\mathrm{v}} = (\tau_1^{\mathrm{v}}, \ldots, \tau_T^{\mathrm{v}})'$.

We use NIMBLE [87] to fit the proposed models both for the real dataset on safety of ICIs and for simulated datasets, with each fit consisting of four independent Markov chain Monte Carlo (MCMC) chains sampling from the joint posterior distribution. We first sample posterior distributions of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Delta}$ and then use the following equations to compute samples from the posterior distributions of the log odds ratio between treatments i and j, and the marginal event rate of treatment t [44]:

$$
\begin{aligned}
\mathrm{LOR}_{ij} &= \mu_i - \mu_j, \\
p_t = E[p_{kt} | \mu_t, \delta_t] &\approx \left[1 + \exp\left(-\mu_t \Big/ \sqrt{1 + \frac{256}{75\pi^2} \delta_t^2} \right) \right]^{-1}.
\end{aligned}
\tag{4.10}
$$

Convergence of chains was assessed by trace plots, sample autocorrelation, and effective sample size. Finally, we can make statistical inference using posterior medians, and 95% equal-tailed credible intervals (CrIs) calculated from the posterior samples.

4.4 Data analysis

In this section, we apply the six models in Table 4.2 to the motivating example of the ICI data and compare the results. This dataset does not have single-arm studies for treatments 7 (2ICIs) and 8 (ICC), so the model settings are slightly different from those described above; to demonstrate the differences, Figure 4.2 presents the directed acyclic graph (DAG) for the DCP model applied to the ICI data. In this DAG, square nodes represent observed data or fixed quantities, circle nodes with white background are intermediate unknown parameters, and circle nodes with gray background are unknown parameters with pre-specified prior distributions, e.g., $\mu_t \sim N(0, 100^2)$ for $t = 7, 8$, $\mu_t^{\mathrm{s}} \sim N(0, 100^2)$ for $t = 1, \ldots, 6$, $\delta_t \sim U(0, 5)$ for $t = 7, 8$, and $\delta_t^{\mathrm{s}} \sim U(0, 5)$ for $t = 1, \ldots, 6$.

4.4.1 Model comparison

The deviance information criterion (DIC) [51] and the logarithm of the pseudo marginal likelihood (LPML) [88, 89] are two popular criteria for comparing Bayesian models. Because estimates of single-arm studies are much less meaningful, we focus on the NMA part of the joint model. Specifically, as a measure to quantify the model's predictive ability, the LPML can be written as:

$$\begin{aligned} \mathrm{LPML} &= \sum_{k=1}^{K} \log(\mathrm{CPO}_k); \\ \mathrm{CPO}_k &= f(\mathcal{D}_k | \mathcal{D}_{-k}, \mathcal{D}_{1:J}^{\mathrm{s}}), \end{aligned} \tag{4.11}$$

where CPO_k is the conditional predictive ordinate of the k^{th} RCT based on the remaining $\mathcal{D}_{-k} = \{\mathcal{D}_l : l \neq k\}$ and on the full data from the single-arm studies $\mathcal{D}_{1:J}^{\mathrm{s}}$. Let $\boldsymbol{\psi}$ represent all unknown parameters $\boldsymbol{\mu}$, $\boldsymbol{\Delta}$, $\boldsymbol{\mu}^{\mathrm{s}}$, $\boldsymbol{\delta}^{\mathrm{s}}$, ρ, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$, $\boldsymbol{\theta}^{\mathrm{s}}$, $\boldsymbol{\iota}^{\mathrm{m}}$, $\boldsymbol{\tau}^{\mathrm{m}}$, $\boldsymbol{\iota}^{\mathrm{v}}$, and $\boldsymbol{\tau}^{\mathrm{v}}$; also, let $\boldsymbol{\psi}_{-k}$ represent all the unknown parameters except $\boldsymbol{\theta}_k$. Then, CPO_k can be

written, using the following identity, as:

$$
\begin{aligned}
& E\left[\frac{1}{f(\mathcal{D}_k|\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k})}\right] \\
&= \int \frac{\pi(\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k}|\mathcal{D}_{1:K},\mathcal{D}^{\mathrm{s}}_{1:J})}{f(\mathcal{D}_k|\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k})} d\boldsymbol{\theta}_k d\boldsymbol{\psi}_{-k} \\
&= \int \frac{f(\mathcal{D}_{1:K},\mathcal{D}^{\mathrm{s}}_{1:J}|\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k})\pi(\boldsymbol{\theta}_k)\pi(\boldsymbol{\psi}_{-k})}{f(\mathcal{D}_k|\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k})f(\mathcal{D}_{1:K},\mathcal{D}^{\mathrm{s}}_{1:J})} d\boldsymbol{\theta}_k d\boldsymbol{\psi}_{-k} \\
&= \int \frac{f(\mathcal{D}_{-k},\mathcal{D}^{\mathrm{s}}_{1:J}|\boldsymbol{\psi}_{-k})f(\mathcal{D}_k|\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k})\pi(\boldsymbol{\theta}_k)\pi(\boldsymbol{\psi}_{-k})}{f(\mathcal{D}_k|\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k})f(\mathcal{D}_{1:K},\mathcal{D}^{\mathrm{s}}_{1:J})} d\boldsymbol{\theta}_k d\boldsymbol{\psi}_{-k} \\
&= \int \frac{f(\mathcal{D}_{-k},\mathcal{D}^{\mathrm{s}}_{1:J}|\boldsymbol{\psi}_{-k})\pi(\boldsymbol{\psi}_{-k})}{f(\mathcal{D}_{1:K},\mathcal{D}^{\mathrm{s}}_{1:J})} d\boldsymbol{\psi}_{-k} \\
&= \frac{f(\mathcal{D}_{-k},\mathcal{D}^{\mathrm{s}}_{1:J})}{f(\mathcal{D}_{1:K},\mathcal{D}^{\mathrm{s}}_{1:J})} = \frac{1}{f(\mathcal{D}_k|\mathcal{D}_{-k},\mathcal{D}^{\mathrm{s}}_{1:J})} = \frac{1}{\mathrm{CPO}_k},
\end{aligned}
\tag{4.12}
$$

where $\pi(\boldsymbol{\theta}_k,\boldsymbol{\psi}_{-k}|\mathcal{D}_{1:K},\mathcal{D}^{\mathrm{s}}_{1:J})$ is the joint posterior distribution in Equation (4.9). Let $\{\boldsymbol{\psi}_c, c=1,\ldots,C\}$ denote the MCMC samples of $\boldsymbol{\psi}$ from this joint posterior distribution; then, CPO_k can be estimated by

$$
\mathrm{CPO}_k \approx \left(\frac{1}{C}\sum_{c=1}^{C}\frac{1}{f(\mathcal{D}_k|\boldsymbol{\psi}_c)}\right)^{-1}, \tag{4.13}
$$

where c indexes MCMC iterations. The DIC can be calculated easily following the steps in Dias et al. [23]. A larger DIC value is less favorable, while larger values of LPML are more favorable. We use the rule of thumb that only differences larger than 5 in DIC indicate a considerable improvement [68].

4.4.2 Treatment ranking

To summarize rankings of the treatments in terms of safety, we use the surface under the cumulative ranking (SUCRA) proposed by Salanti et al. [66]. Let prob_{ti} be the probability that treatment t has the i^{th} rank, where $i=1$ represents the safest treatment; the SUCRA of the t^{th} treatment is

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

where posterior mean of prob_{ti} is easily calculated using MCMC samples. The SUCRA ranges from 0% to 100%; a higher SUCRA value implies a better treatment.

4.4.3 Results

Table 4.3 presents the results for absolute risk of events for the t^{th} treatment (p_t), fixed effect of log-odds for the t^{th} treatment (μ_t), standard deviation of log-odds for the t^{th} treatment (δ_t), selected log odds ratios LOR_{ij}, LPML, and DIC. These models did not differ notably in LPML and DIC. However, some differences appear in the estimates and intervals.

Figure 4.3A presents a forest plot of the standard deviations δ_t. Clearly, because of lack of information about treatments 3 (IPI high), 5 (ATE), and 7 (2ICIs), the estimates of δ_3, δ_5, and δ_7 under the NB method were dominated by prior information, i.e., $U(0, 5)$, with wide CrIs. By fully (the FBV method) or adaptively (the CPV method) incorporating variance information from single-arm studies for treatments IPI high and ATE, we may have better estimates for δ_3 and δ_5 with much narrower CrIs. However, when the RCTs provided a good deal of information, e.g., for treatment 6 (ICI+ICC with $B_6 = 6 > 5$) and the variances in the RCTs and single-arm studies differed, the FBV method had a much stronger effect on the posterior of δ_6 than the adaptive (CPV) method; the posterior median and 95% CrI of δ_6 were 0.18 (0.04, 0.57) for NB, 0.21 (0.04, 0.66) for CPV, and 0.35 (0.10, 0.91) for FBV. Similar results were obtained when mean information was adaptively or fully borrowed, e.g., the posterior median and 95% CrI of μ_1 were -1.84 (-2.19, -1.58) for NB, -1.75 (-2.08, -1.51) for CPM, and -1.71 (-1.90, -1.52) for FBMV. Posterior medians of the μ_t's were generally quite similar among the NB, CPV, and FBV methods because the CPV and FBV methods shared only variance information. The CPM method also narrowed the CrIs of δ_3 and δ_5 a bit by sharing mean information from single-arm studies.

The differences between methods in mean and variance estimates can affect the estimates of absolute risks, as shown in Figure 4.3B. The NB method yielded a much wider CrI for treatments IPI high dose and ATE than the CPV and FBV methods because of those treatments had few RCTs. On the other hand, the FBV method gave wide CrIs for treatment 6 (ICI+ICC) because it fully incorporated variance information from the single-arm studies, while the CPV method gave estimates more similar to the NB method by adaptively downweighting variance information that was inconsistent between the RCTs and single-arm studies; the posterior median and 95% CrI of p_6 were 0.47 (0.42, 0.53) for NB, 0.47 (0.41, 0.54) for CPV, and 0.48 (0.39, 0.57) for FBV.

The CPV and FBV methods provided almost the same posterior medians of p_t for all treatments as the NB method, while the other three methods (CPM, DCP, and FBMV) gave rather different point estimates of p_t because of incorporating potentially inconsistent results from single-arm studies; the posterior median and 95% CrI of p_2 were 0.19 (0.13, 0.27) for NB, 0.19 (0.14, 0.27) for CPV, 0.20 (0.14, 0.28) for CPM, and 0.21 (0.15, 0.28) for FBMV.

Figure 4.4 visualizes the estimated log odds ratios (the off-diagonal part) and SUCRAs (the diagonal part, shown in the percent) under the NB (upper right) and CPV (lower left) methods. Specifically, the radius of the gray circle represents the point estimate of LOR_{ij}, with the radius of the inner white circle (not shown if $P > 0.05$ for testing the difference between two treatments) and outer colored circle representing the 95% CrI. The coloration on the scale is determined by the P-value, with blue indicating that the upper-left treatment is better than the lower-right treatment in terms of lower drug-related grade 3–5 adverse events (AEs). The largest difference between NB and CPV methods in estimating LOR_{ij} and SUCRA_t was for the log odds ratio between ATE and IPI high dose; the posterior median and 95% CrI of LOR_{53} were -1.25 (-3.20, 0.23) for NB and -1.17 (-2.01, -0.40) for CPV. Such differences arose because the CPV method incorporated more variance information than the NB method, which narrowed the CrI.

These analyses confirmed that drug-related grade 3–5 AEs were dose-dependent with ipilimumab; the posterior median and 95% CrI of LOR_{23} were -0.97 (-1.90, -0.11) for NB, and -0.95 (-1.63, -0.30) for CPV. Also, drug-related grade 3–5 AEs were less frequent for all ICIs (NIV, PEM, ATE, and IPI low) than for traditional chemotherapy or combination therapy of ICI and ICC. We found no significant differences between anti-PD-1 monotherapy (nivolumab, or pembrolizumab), anti-PD-L1 monotherapy (atezolizumab), and anti-CTLA-4 monotherapy (ipilimumab 3mg/kg every 3 weeks) in drug-related grade 3–5 AEs, with appropriate dose. Based on SUCRA, however, nivolumab ($\text{SUCRA}_1 = 0.92$ for the CPV method) may be the ICI drug with the lowest frequency of drug-related grade 3–5 AEs among the drugs that were investigated.

In summary, when results from single-arm studies were potentially unreliable, the CPV method can provide better estimates than other methods that borrow mean information. At the same time, unlike the NB or FBV methods, the CPV method allows

treatments with limited data to adaptively incorporate variance information from single-arm studies to improve estimates related to these treatments.

4.5 Simulation studies

4.5.1 Simulation settings

These simulation studies compared five methods (NB, CPV, FBV, CPM, and DCP) defined in Section 4.3. Each simulated dataset contained $K = 14$ RCTs, $J = 30$ single-arm studies, and $T = 5$ treatments (indexed from 1 to 5). The number of participants in each treatment arm in each RCT, n_{kt}, was fixed at 150. The 30 single-arm studies were allocated to treatments with the pre-specified partition scheme of $B_1^{\mathrm{s}} = 14$, $B_2^{\mathrm{s}} = 5$, $B_3^{\mathrm{s}} = 4$, $B_4^{\mathrm{s}} = 4$, and $B_5^{\mathrm{s}} = 3$. The number of participants in each single-arm study was 75, 38, 75, 150, and 113 for treatments 1 to 5 respectively. The number of simulated datasets in each simulation setting was 1000.

In simulated datasets, we considered four scenarios with different levels of reliability of mean and variance information from single-arm studies. In scenario EM-EV (equal mean and equal variance), we first generated a complete dataset for the RCTs under the AB model with binary outcomes as in Equation (4.1), with $\boldsymbol{\mu}^{\mathrm{n}} = (\mu_1, \ldots, \mu_5)' = (-2, -3, -2.5, -2, -1.5)'$ and $(\theta_{k1}, \ldots, \theta_{k5})' \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Delta}\mathbf{P}\boldsymbol{\Delta}$, the correlation matrix $\mathbf{P}$ had all off-diagonal entries $\rho_{ij} = 0.5$ for $i \neq j$, and standard deviations (i.e., diagonal entries) $\boldsymbol{\delta}^{\mathrm{n}} = (\delta_1, \ldots, \delta_5)' = (0.4, 1.0, 1.0, 0.3, 0.3)'$. In this complete dataset for the RCTs, each trial had 5 arms with $|\mathcal{A}_k| = 5$. To generate the 30 single-arm studies, we used Equation (4.2) with $\boldsymbol{\mu}^{\mathrm{s}} = (\mu_1^{\mathrm{s}}, \ldots, \mu_5^{\mathrm{s}})' = (-2, -3, -2.5, -2, -1.5)'$, and $\boldsymbol{\delta}^{\mathrm{s}} = (\delta_1^{\mathrm{s}}, \ldots, \delta_5^{\mathrm{s}})' = (0.4, 1.0, 1.0, 0.3, 0.3)'$. In these single-arm studies, each study only had one arm with $|\mathcal{A}_j^{\mathrm{s}}| = 1$. The only difference between EM-EV and the other three scenarios was the pre-specified values for $\boldsymbol{\mu}^{\mathrm{n}}$, $\boldsymbol{\delta}^{\mathrm{n}}$, $\boldsymbol{\mu}^{\mathrm{s}}$, and $\boldsymbol{\delta}^{\mathrm{s}}$, as follows: 1) UM-EV (unequal mean and equal variance) scenario had $\boldsymbol{\mu}^{\mathrm{n}} = (-2, -3, -2.5, -2, -1.5)'$, $\boldsymbol{\mu}^{\mathrm{s}} = (-1, -2.5, -2, -2.5, -2)'$, and $\boldsymbol{\delta}^{\mathrm{n}} = \boldsymbol{\delta}^{\mathrm{s}} = (0.4, 1.0, 1.0, 0.3, 0.3)'$; 2) EM-UV (equal mean and unequal variance) scenario had $\boldsymbol{\mu}^{\mathrm{n}} = \boldsymbol{\mu}^{\mathrm{s}} = (-2, -3, -2.5, -2, -1.5)'$, $\boldsymbol{\delta}^{\mathrm{n}} = (0.4, 1.0, 1.0, 0.3, 0.3)'$, and $\boldsymbol{\delta}^{\mathrm{s}} = (1.2, 0.5, 0.5, 0.9, 0.9)'$; 3) UM-UV (unequal mean and unequal variance) scenario had $\boldsymbol{\mu}^{\mathrm{n}} = (-2, -3, -2.5, -2, -1.5)'$, $\boldsymbol{\mu}^{\mathrm{s}} = (-1, -2.5, -2, -2.5, -2)'$, $\boldsymbol{\delta}^{\mathrm{n}} = (0.4, 1.0, 1.0, 0.3, 0.3)'$,

and $\boldsymbol{\delta}^{\mathrm{s}} = (1.2, 0.5, 0.5, 0.9, 0.9)'$.

Once the complete RCTs dataset was generated, we excluded treatment arms to create a realistic (partially missing) NMA dataset as illustrated in Figure 4.5, with two types of missingness: 1) missing completely at random (MCAR) and 2) missing at random (MAR). Under the MCAR mechanism, we kept all treatment 1 data (all 14 RCTs) and then randomly kept data for treatments 2 to 5 data in blocks of 3, 4, 2, and 5 trials respectively, where the blocks did not overlap. Under the MAR mechanism, we kept all treatment 1 data and ranked the RCTs in descending order by the rough estimates of event rates r_{k1}/n_{k1}; then we made treatment 5 available only in the first 5 trials, treatment 4 available in next 2, and so on as in Figure 4.5.

We used the prior specifications in Table 4.2 for all models and obtained posterior medians and 95% equal-tailed CrIs for these estimands: event risk for the t^{th} treatment (p_t), fixed effect of treatment-specific log-odds (μ_t), standard deviation of treatment-specific log-odds (δ_t), and log odds ratio between the i^{th} and j^{th} treatments (LOR_{ij}). To measure the methods' performance, we used bias, mean squared error (MSE), and the 95% CrI's coverage probability (CP) and length (CrIL).

4.5.2 Simulation results

Table 4.4 summarizes the bias and MSE of the posterior median and the coverage probability of the 95% CrI for the five methods in four different simulation scenarios under MAR. Due to space limits, instead of presenting the results for each treatment or each treatment comparison, for each of bias, MSE and CP, we summarized overall measures across all treatments or across all pairs of comparisons. For example, the entry with bias as the column and LOR_{ij} as the row was calculated as $\sum_{i \neq j} |\text{bias}(\text{LOR}_{ij})|$. The formula was similar for MSE: $\sum_{i \neq j} \text{MSE}(\text{LOR}_{ij})$. To summarize the CPs, the corresponding value in column CP and row p_t was calculated as $\sum_{t=1}^{5} (0.95 - \text{CP}(p_t))_+$, where $x_+ = x$ if $x \geq 0$ and $x_+ = 0$ if $x < 0$. Table 4.5 presents the simulation results under MCAR with similar summaries. Figure 4.6 displays the log of CrIL ratio for four methods (NB, FBV, CPM, and DCP) compared to the CPV method, using box plots with whiskers representing the 1st and 99th percentiles. Each sub-figure presents one of the four estimands: μ_t, δ_t, p_t, and LOR_{ij}. Also, each panel shows one of the four scenarios (EM-EV, UM-EV, EM-UV, and UM-UV in columns) under the different

missingness structures (MCAR and MAR in rows).

Under the MAR mechanism (Table 4.4), the CPV method was much better than the NB method with less biased estimates, smaller MSE, and comparable CP in all scenarios. Although the CPV method produced the second largest MSE (smaller only than that of the NB method), the other three methods did not perform well in some situations. For example, DCP was better than CPV in terms of bias when mean information from single-arm studies was reliable (EM-UV and EM-EV); however, when this information was not reliable (UM-EV and UM-UV), the biases of DCP and CPM were even worse than that of NB. Similarly, FBV's performance was worse than CPV's when variance information from single-arm studies was not reliable (EM-UV and UM-UV). Similar performance patterns were present under the MCAR mechanism (Table 4.5), though the difference between methods was much smaller.

Compared to the NB method, the CPV and FBV methods greatly reduced CrI length for all estimates under all scenarios (Figure 4.6). Compared to CPV, the relative length of CrI for CPM varied depending on simulation scenario and estimand. The DCP method produced the smallest CrI length for all estimates under all scenarios; hence, when we believe in the reliability of mean information from single-arm studies, the DCP method would be the first choice.

Overall, the CPV method provided better estimates of log odds ratios and absolute risks than the NB method even when true variances in the single-arm studies differed from true variances in the RCTs, e.g., the elementwise ratios were $\boldsymbol{\delta}^{\mathrm{n}}/\boldsymbol{\delta}^{\mathrm{s}} = (1/3, 2, 2, 1/3, 1/3)'$. The performance of the other three borrowing strategies depended largely on the reliability of the single-arm studies.

4.6 Summary and discussion

This chapter has proposed and discussed different strategies to incorporate single-arm studies into arm-based NMA to mitigate the prevalent "lack of information" problem. We have performed extensive simulation studies to explore whether it is preferable to choose a fully borrowing strategy or one of the adaptively borrowing methods (commensurate prior), and whether to borrow mean information, variance information, or both. The simulation studies considered scenarios when information from single-arm studies

could be unreliable. Our proposed CPV method delivered the most robust estimates of relative and absolute risks in all four simulation scenarios. Specifically, CPV could gain efficiency even in the presence of modestly discordant variance information, by facilitating partial pooling of variance information from single-arm studies, rather than fully borrowing as in the FBV method. Also, unlike the CPM and DCP methods, ignoring mean information from the supplemental source could help the CPV method produce more reliable point estimates with reduced CrI lengths. As far as we know, this is the first proposal in the field of Bayesian extrapolation analyses to borrow only variance information. The application to safety of ICIs in cancer research also illustrated potential gains of the CPV method for estimates related to the treatments IPI high dose and ATE, by adaptively incorporating variance information from single-arm studies.

We have focused on commensurate priors to synthesize RCTs and single-arm studies in the AB-NMA; many future studies are possible. First, selecting pre-specified values for p^{v}, R^{v}, $s_{\mathrm{l}}^{\mathrm{v}}$, and $s_{\mathrm{u}}^{\mathrm{v}}$ in the commensurate prior might lead to some problems [76]; future research is needed specifically for CPV in the AB-NMA. Also, existing methods cannot assess the importance of single-arm studies in the AB-NMA. We need to develop methods that separately assess each component of the joint model, of the NMA dataset and the supplemental source, e.g., by decomposing DIC or LPML into two parts [12], with one part for the supplemental source and the other part for the NMA dataset conditional on the supplemental source.

Second, empirical studies should evaluate the reliability of mean and especially variance information from single-arm studies that may be eligible for use in NMAs. Such a study could help experts to judge whether incorporating variance is reasonable in general or in specific subject-matter areas.

Third, alternative Bayesian methods could be used to adaptively incorporate information from single-arm studies into the AB-NMA. For example, the extrapolation strategy for meta-analyses proposed by Röver et al. [35] is to express the posterior distribution as a weighted average of posterior components from four simple models: NB, FBM, FBV, and FBMV. Although not mentioned by Röver et al. [35], variance information only could be borrowed by averaging just two models: NB and FBV. Another method is the power priors [34,73] that incorporate supplemental information by raising the likelihood of the single-arm studies to a power $\alpha \in [0, 1]$. This may be problematic,

however, because mean and variance information is included in the likelihood as a whole and cannot easily be separated, as in the CPV and CPM methods or in Röver et al. [35].

Finally, Turner et al. [37] proposed to incorporate external evidence in the CB-NMA by using informative priors specified using previously published evidence for describing between-trial heterogeneity [60]. A similar idea could be used in the AB-NMA by first obtaining a posterior distribution for δ_t^{s} from single-arm studies, then using this as an informative prior for δ_t. However, small B_t^{s} or B_t may still provide litter improvement for the estimation.

Table 4.1: Safety of ICIs on cancer data set. Study index, reference of each study, treatment details, number of treatment-related grade 3–5 adverse events (r), number of patients assigned in each treatment arm (n) are presented. ICI=immune checkpoint inhibitor; NIV=nivolumab; IPI=ipilimumab; PEM=pembrolizumab; ATE=atezolizumab; ICC=investigator's choice chemotherapy.

Study	Reference	Treatment	r	n
Phase II or III RCT				
1	[90]	NIV: 3mg/kg every 2 weeks	65	452
1		IPI: 10mg/kg every 3 weeks	210	453
2	[91]	NIV: 3mg/kg every 2 weeks	68	313
2	& [92]	2ICIs: NIV+IPI	186	311
2		IPI: 3mg/kg every 3 weeks	87	313
3	[93]	PEM: 10mg/kg every 2 weeks	48	278
3	& [94]	PEM: 10mg/kg every 3 weeks	46	277
3		IPI: 3mg/kg every 3 weeks	50	256
4	[95]	ATE: 1200mg every 3 weeks	90	609
4		ICC	248	578
5	[96]	NIV: 3mg/kg every 2 weeks	37	268
5	& [97]	ICC	35	102
6	[98]	ICI+ICC	103	247
6	& [99]	ICC	15	251
7	[100]	NIV: 3mg/kg every 2 weeks	32	236
7		ICC	40	111
8	[101]	ICI+ICC	205	388
8		ICC	129	361
9	[102]	NIV: 3mg/kg every 2 weeks	76	406
9		ICC	147	397
10	[103]	NIV: 3mg/kg every 3 weeks	49	267
10		ICC	136	263
11	[104]	PEM: 200mg every 3 weeks	40	266
11		ICC	126	255
12	[105]	IPI: 10mg/kg every 3 weeks	128	364
12		IPI: 3mg/kg every 3 weeks	68	362
13	[106]	PEM: 200mg every 3 weeks	41	154
13		ICC	80	150
14	[107]	ICI+ICC	231	478
14		ICC	214	476
15	[108]	ICI+ICC	23	59
15		ICC	16	62
16	[109]	2ICIs: NIV+IPI	54	94
16	& [110]	IPI: 3mg/kg every 3 weeks	9	46
17	[111]	PEM: 2 mg/kg every 3 weeks	43	339
17		PEM: 10 mg/kg every 3 weeks	55	343
17		ICC	114	309
18	[112]	ATE: 1200mg every 3 weeks	17	142
18		ICC	55	135
19	[113]	NIV: 3mg/kg every 2 weeks	24	206
19		ICC	36	205
20	[114]	ICC	45	171
20		PEM: 2 mg/kg every 3 weeks	19	178
20		PEM: 10 mg/kg every 3 weeks	25	179
21	[115]	NIV: 3mg/kg every 2 weeks	9	131
21		ICC	75	129
22	[116]	NIV: 3mg/kg every 2 weeks	31	287
22		ICC	145	268
23	[117]	ATE: 1200mg every 3 weeks	95	459
23		ICC	198	443
24	[118]	ICI+ICC	40	84
24		ICC	13	44
25	[119]	ICC	25	65
25		ICI+ICC	56	138
26	[120]	IPI: 3mg/kg every 3 weeks	7	40
26		IPI: 10mg/kg every 3 weeks	14	42
27	[121]	IPI: 3mg/kg every 3 weeks	6	71
27		IPI: 10mg/kg every 3 weeks	18	71
Single arm trials				
1	[122]	PEM: 200mg every 3 weeks	6	40
2	[123]	PEM: 10mg/kg every 2 weeks	5	36
3	[124]	IPI: 10mg/kg every 3 weeks	9	25
4	[125]	PEM: 2mg/kg every 3 weeks	4	26
5	[126]	NIV: 3mg/kg every 2 weeks	20	80
6	[127]	ICI+ICC	30	46
7	[128]	ICI+ICC	47	86
8	[129]	NIV: 3mg/kg every 2 weeks	4	10
9	[130]	NIV: 3mg/kg every 2 weeks	2	35
10	[131]	NIV: 3mg/kg every 2 weeks	11	65
11	[132]	NIV: 3mg/kg every 2 weeks	4	17
12	[133]	NIV: 3mg/kg every 2 weeks	17	76
13	[134]	NIV: 3mg/kg every 2 weeks	15	74
14	[135]	NIV: 3mg/kg every 2 weeks	51	270
15	[136]	IPI: 3mg/kg every 3 weeks	3	20
16	[137]	NIV: 3mg/kg every 2 weeks	3	23
17	[138]	ICI+ICC	11	15
18	[139]	IPI: 3mg/kg every 3 weeks	20	53
19	[140]	IPI: 3mg/kg every 3 weeks	20	103
20	[141]	ATE: 1200mg every 3 weeks	20	119
21	[142]	PEM: 200mg every 3 weeks	26	171
22	[143]	IPI: 10mg/kg every 3 weeks	39	155
23	[144]	ATE: 1200mg every 3 weeks	82	659
24	[145]	NIV: 3mg/kg every 2 weeks	22	117
25	[146]	ATE: 1200mg every 3 weeks	50	310
26	[147]	NIV: 3mg/kg every 2 weeks	39	330
27	[148]	IPI: 10mg/kg every 3 weeks	145	393
28	[86]	ICI+ICC	9	35

Figure 4.1: Network plot of the dataset about safety of ICIs in treating cancer. Each node represents a regimen, and each edge represents a direct comparison between two regimens. Vertex radius is proportional to the number of RCTs containing the regimen (dark inner circle) plus the number of single-arm studies of the regimen (light outer circle); edge thickness is proportional to the number of direct comparisons. Numbers in parentheses under a regimen name include the number of RCTs and the number of single-arm studies that investigate the regimen (e.g., 9 RCTs and 11 single-arm studies investigate nivolumab).

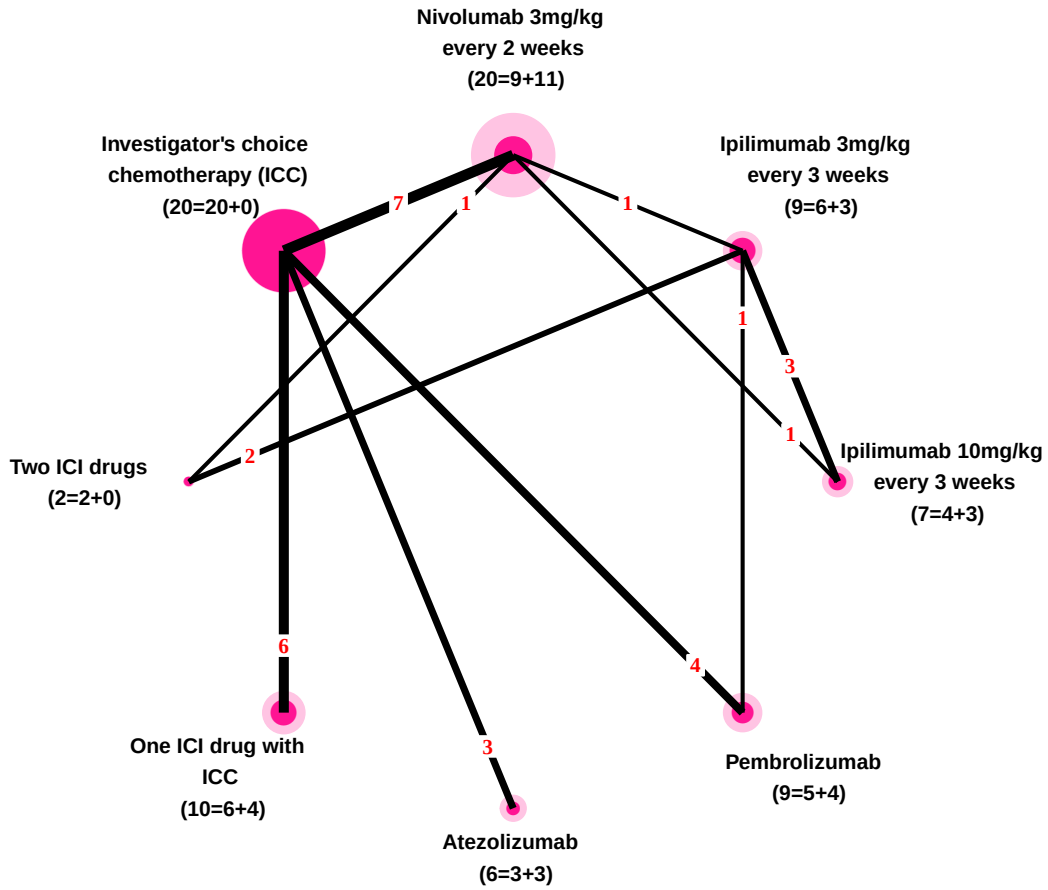


Table 4.2: Summary of prior specifications and assumptions for μ_t, δ_t, μ_t^{s}, δ_t^{s}, and the correlation matrix $\mathbf{P}$, for six different models.

Model	Parameter				
	μ_t, $t = 1, \ldots, T$	δ_t, $t = 1, \ldots, T$	μ_t^{s}, $t = 1, \ldots, T$	δ_t^{s}, $t = 1, \ldots, T$	$\mathbf{P} = \{\rho_{ij}\}$
NB	$\mu_t \sim N(0, 100^2)$	$\delta_t \sim U(0, 5)$	NA	NA	$\rho_{ij} = \rho$ $(i \neq j)$ and $\rho \sim U(-\frac{1}{T-1}, 1)$ for all models
FBMV	$\mu_t \sim N(0, 100^2)$	$\delta_t \sim U(0, 5)$	$\mu_t = \mu_t^{\mathrm{s}}$	$\delta_t = \delta_t^{\mathrm{s}}$	
FBV	$\mu_t \sim N(0, 100^2)$	$\delta_t \sim U(0, 5)$	$\mu_t^{\mathrm{s}} \sim N(0, 100^2)$	$\delta_t = \delta_t^{\mathrm{s}}$	
CPM	Equation (4.4) with $p^{\mathrm{m}} = 0.5$, $R^{\mathrm{m}} = 2500$, $s_{\mathrm{l}}^{\mathrm{m}} = 0$, and $s_{\mathrm{u}}^{\mathrm{m}} = 2$	$\delta_t \sim U(0, 5)$	$\mu_t^{\mathrm{s}} \sim N(0, 100^2)$	$\delta_t^{\mathrm{s}} \sim U(0, 5)$	
CPV	$\mu_t \sim N(0, 100^2)$	Equation (4.6) with $p^{\mathrm{v}} = 0.5$, $R^{\mathrm{v}} = 2500$, $s_{\mathrm{l}}^{\mathrm{v}} = 0$, and $s_{\mathrm{u}}^{\mathrm{v}} = 2$	$\mu_t^{\mathrm{s}} \sim N(0, 100^2)$	$\delta_t^{\mathrm{s}} \sim U(0, 5)$	
DCP	Equation (4.4) with $p^{\mathrm{m}} = 0.5$, $R^{\mathrm{m}} = 2500$, $s_{\mathrm{l}}^{\mathrm{m}} = 0$, and $s_{\mathrm{u}}^{\mathrm{m}} = 2$	Equation (4.6) with $p^{\mathrm{v}} = 0.5$, $R^{\mathrm{v}} = 2500$, $s_{\mathrm{l}}^{\mathrm{v}} = 0$, and $s_{\mathrm{u}}^{\mathrm{v}} = 2$	$\mu_t^{\mathrm{s}} \sim N(0, 100^2)$	$\delta_t^{\mathrm{s}} \sim U(0, 5)$	

Figure 4.2: Directed acyclic graph of the DCP model for the motivating example. □, observed data or fixed quantities; ◯, intermediate unknown parameters; ●, unknown parameters with pre-specified prior distributions.

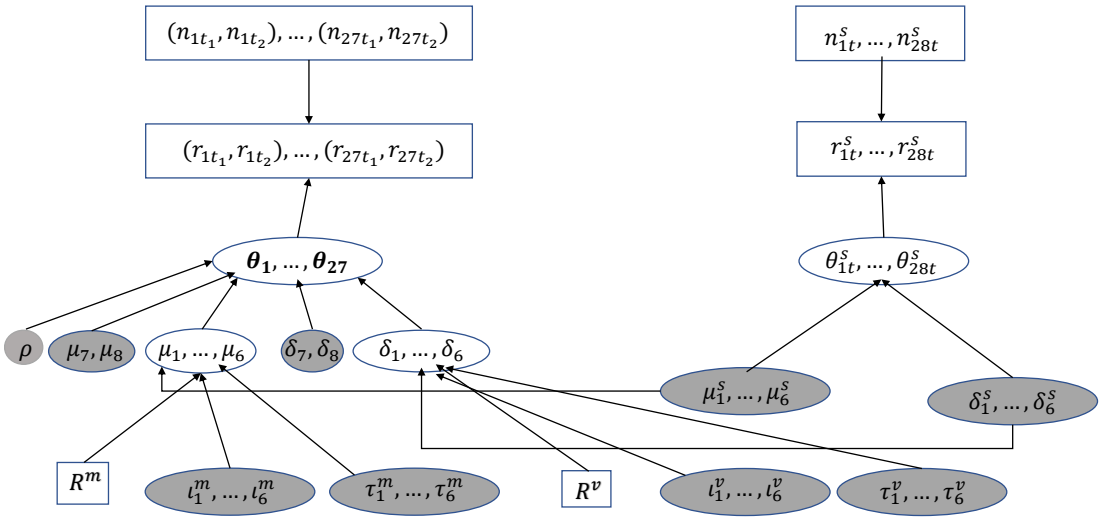
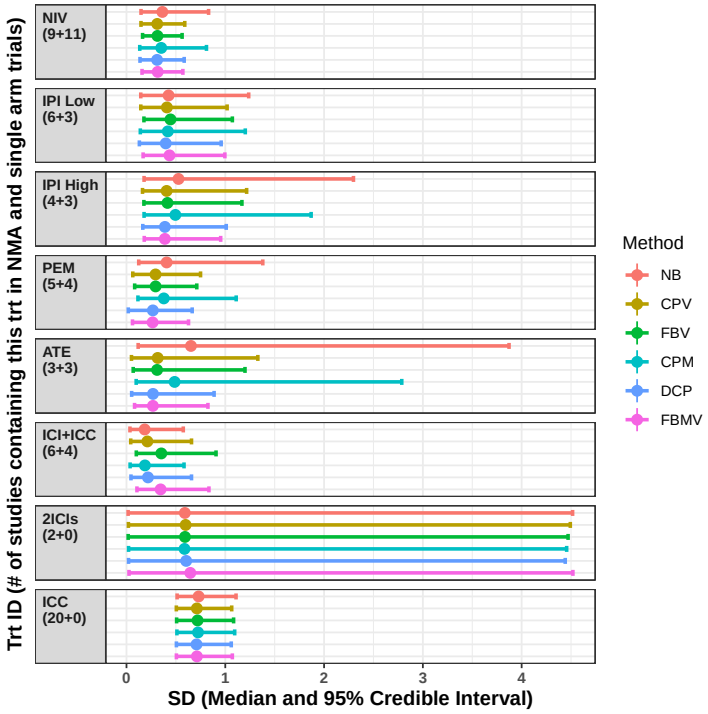


Table 4.3: Analysis of safety of ICIs in cancer treatment: comparison of posterior medians and 95% credible intervals for 6 different models (NB, CPV, FBV, CPM, DCP, and FBMV), specifically absolute risk of events for the t^{th} treatment (p_t), fixed effect of log-odds for the t^{th} treatment (μ_t), standard deviation of the log-odds for the t^{th} treatment (δ_t), and log odds ratio LOR_{ij} comparing the i^{th} and j^{th} treatments. Treatment labels: 1) NIV; 2) IPI low; 3) IPI high; 4) PEM; 5) ATE; 6) ICI+ICC; 7) 2ICIs; and 8) ICC.

Parameter	Posterior median (95% credible interval)					
	NB	CPV	FBV	CPM	DCP	FBMV
p_1	0.14 (0.11, 0.18)	0.14 (0.12, 0.17)	0.14 (0.12, 0.17)	0.15 (0.12, 0.19)	0.15 (0.12, 0.18)	0.16 (0.14, 0.19)
p_2	0.19 (0.13, 0.27)	0.19 (0.14, 0.27)	0.19 (0.13, 0.27)	0.20 (0.14, 0.28)	0.20 (0.15, 0.27)	0.21 (0.15, 0.28)
p_3	0.38 (0.25, 0.56)	0.37 (0.27, 0.50)	0.37 (0.27, 0.50)	0.36 (0.27, 0.49)	0.36 (0.28, 0.47)	0.36 (0.28, 0.45)
p_4	0.16 (0.12, 0.26)	0.16 (0.13, 0.22)	0.16 (0.13, 0.22)	0.16 (0.13, 0.22)	0.16 (0.13, 0.20)	0.16 (0.13, 0.20)
p_5	0.16 (0.09, 0.41)	0.16 (0.11, 0.25)	0.16 (0.11, 0.24)	0.16 (0.12, 0.30)	0.16 (0.12, 0.21)	0.16 (0.12, 0.21)
p_6	0.47 (0.42, 0.53)	0.47 (0.41, 0.54)	0.48 (0.39, 0.57)	0.48 (0.42, 0.54)	0.48 (0.41, 0.55)	0.49 (0.42, 0.57)
p_7	0.54 (0.14, 0.73)	0.54 (0.14, 0.75)	0.54 (0.16, 0.75)	0.55 (0.16, 0.75)	0.55 (0.16, 0.75)	0.55 (0.17, 0.78)
p_8	0.38 (0.31, 0.46)	0.38 (0.31, 0.46)	0.38 (0.31, 0.46)	0.39 (0.32, 0.46)	0.39 (0.32, 0.46)	0.39 (0.32, 0.47)
μ_1	−1.84 (−2.19, −1.58)	−1.82 (−2.11, −1.60)	−1.82 (−2.09, −1.59)	−1.75 (−2.08, −1.51)	−1.74 (−2.03, −1.54)	−1.71 (−1.90, −1.52)
μ_2	−1.48 (−2.06, −1.07)	−1.48 (−1.99, −1.08)	−1.49 (−2.03, −1.06)	−1.44 (−1.93, −1.05)	−1.42 (−1.87, −1.06)	−1.40 (−1.82, −1.04)
μ_3	−0.53 (−1.36, 0.33)	−0.53 (−1.10, 0.01)	−0.54 (−1.10, −0.01)	−0.59 (−1.11, −0.03)	−0.58 (−1.00, −0.14)	−0.60 (−0.99, −0.22)
μ_4	−1.67 (−2.18, −1.22)	−1.67 (−2.00, −1.34)	−1.66 (−1.99, −1.34)	−1.69 (−2.02, −1.38)	−1.68 (−1.94, −1.42)	−1.67 (−1.93, −1.43)
μ_5	−1.75 (−3.52, −0.59)	−1.69 (−2.36, −1.22)	−1.69 (−2.31, −1.23)	−1.75 (−2.33, −1.25)	−1.71 (−2.09, −1.40)	−1.72 (−2.06, −1.42)
μ_6	−0.10 (−0.34, 0.14)	−0.11 (−0.38, 0.17)	−0.10 (−0.47, 0.32)	−0.09 (−0.31, 0.16)	−0.09 (−0.36, 0.19)	−0.04 (−0.32, 0.30)
μ_7	0.16 (−4.09, 1.59)	0.18 (−3.84, 1.86)	0.19 (−3.55, 1.92)	0.20 (−3.41, 1.84)	0.21 (−3.37, 1.85)	0.24 (−3.36, 2.23)
μ_8	−0.52 (−0.86, −0.18)	−0.52 (−0.86, −0.19)	−0.52 (−0.86, −0.18)	−0.50 (−0.83, −0.17)	−0.51 (−0.84, −0.17)	−0.49 (−0.82, −0.15)
δ_1	0.36 (0.15, 0.83)	0.31 (0.15, 0.59)	0.31 (0.16, 0.56)	0.35 (0.13, 0.81)	0.31 (0.14, 0.58)	0.32 (0.16, 0.57)
δ_2	0.43 (0.14, 1.24)	0.41 (0.14, 1.02)	0.45 (0.18, 1.07)	0.42 (0.14, 1.20)	0.40 (0.13, 0.96)	0.44 (0.17, 1.00)
δ_3	0.53 (0.18, 2.30)	0.41 (0.16, 1.22)	0.41 (0.18, 1.17)	0.49 (0.18, 1.87)	0.39 (0.16, 1.01)	0.39 (0.18, 0.95)
δ_4	0.41 (0.12, 1.38)	0.29 (0.06, 0.75)	0.29 (0.08, 0.71)	0.38 (0.12, 1.11)	0.27 (0.02, 0.66)	0.26 (0.06, 0.63)
δ_5	0.65 (0.12, 3.87)	0.32 (0.05, 1.33)	0.31 (0.07, 1.20)	0.49 (0.10, 2.79)	0.27 (0.05, 0.89)	0.27 (0.08, 0.82)
δ_6	0.18 (0.04, 0.57)	0.21 (0.04, 0.66)	0.35 (0.10, 0.91)	0.19 (0.04, 0.58)	0.22 (0.05, 0.66)	0.35 (0.11, 0.83)
δ_7	0.59 (0.02, 4.52)	0.60 (0.02, 4.49)	0.59 (0.02, 4.47)	0.59 (0.02, 4.46)	0.60 (0.02, 4.44)	0.65 (0.03, 4.52)
δ_8	0.73 (0.51, 1.11)	0.71 (0.50, 1.07)	0.72 (0.51, 1.08)	0.72 (0.51, 1.10)	0.71 (0.51, 1.06)	0.71 (0.51, 1.07)
LOR_{14}	−0.17 (−0.72, 0.38)	−0.16 (−0.57, 0.23)	−0.16 (−0.56, 0.23)	−0.07 (−0.50, 0.33)	−0.07 (−0.44, 0.26)	−0.03 (−0.34, 0.28)
LOR_{15}	−0.08 (−1.26, 1.66)	−0.13 (−0.67, 0.55)	−0.14 (−0.65, 0.51)	−0.01 (−0.57, 0.60)	−0.04 (−0.44, 0.38)	0.01 (−0.34, 0.39)
LOR_{45}	0.09 (−1.18, 1.88)	0.03 (−0.55, 0.75)	0.03 (−0.52, 0.71)	0.06 (−0.52, 0.70)	0.04 (−0.37, 0.48)	0.04 (−0.34, 0.46)
LOR_{23}	−0.97 (−1.90, −0.11)	−0.95 (−1.63, −0.30)	−0.96 (−1.63, −0.30)	−0.87 (−1.53, −0.23)	−0.85 (−1.41, −0.32)	−0.80 (−1.33, −0.29)
LOR_{53}	−1.25 (−3.20, 0.23)	−1.17 (−2.01, −0.40)	−1.15 (−1.97, −0.42)	−1.16 (−1.99, −0.43)	−1.13 (−1.71, −0.61)	−1.11 (−1.63, −0.63)
DIC	15970.4	15971.0	15970.3	15970.7	15971.4	15970.5
LPML	−224.4	−226.3	−222.4	−221.7	−224.3	−226.3

Figure 4.3: Results for the dataset of the safety of ICIs in cancer treatment: forest plot of posterior estimates of standard deviations δ_t and absolute risks p_t (posterior median with 95% credible interval). Different colors indicate different methods. The y-axis represents regimen abbreviations, with $B_t + B_t^{\mathrm{s}}$ in parentheses.

(A) standard deviation δ_t



(B) absolute risk p_t

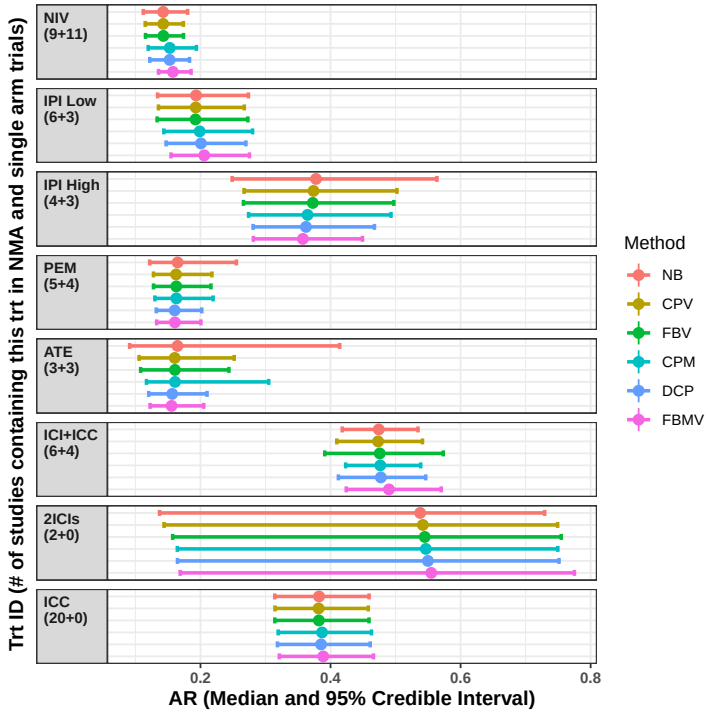


Figure 4.4: Estimated log odds ratios LOR_{ij} for grade 3–5 adverse events of ICIs in cancer patients using the NB (upper right) and CPV (lower left) methods. The LOR information is visualized as a plate plot, with the gray circle representing the posterior median of LOR_{ij} and the inner white circle (not shown if P-value > 0.05) and outer colored circle representing the 95% CrI. The coloration is determined by the P-value of LOR, with blue indicating the upper-left treatment is safer than the lower-right treatment. The diagonal of the plot displays SUCRA of treatments under the NB (upper right number) and CPV (lower left number) methods.

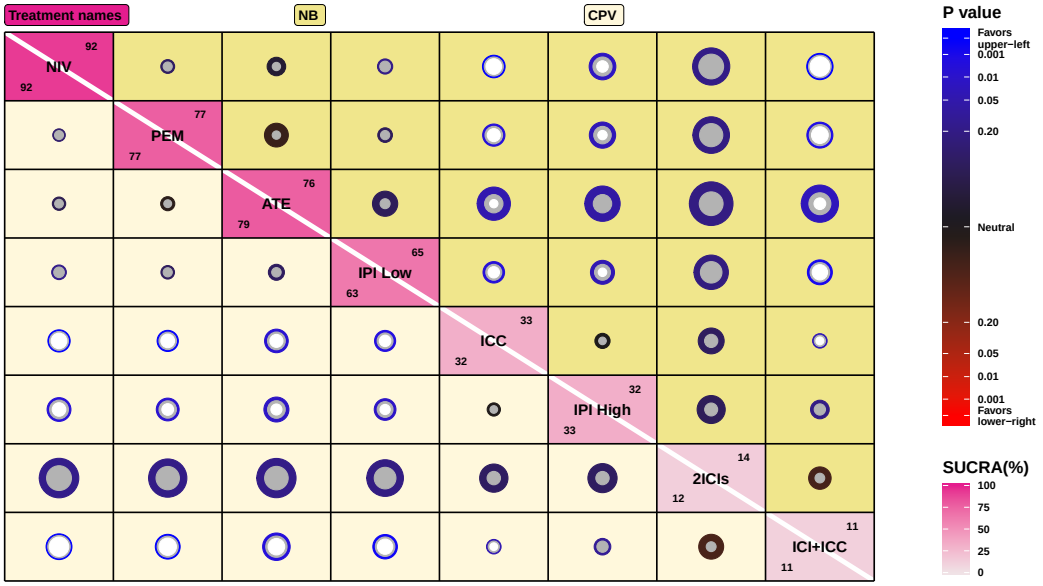


Figure 4.5: Missing data structures for the simulation study under MCAR and MAR. The number in the white-background boxes is the observed number of RCTs for the treatment in the corresponding column, while the gray background indicates that the corresponding treatment is not observed in these trials.

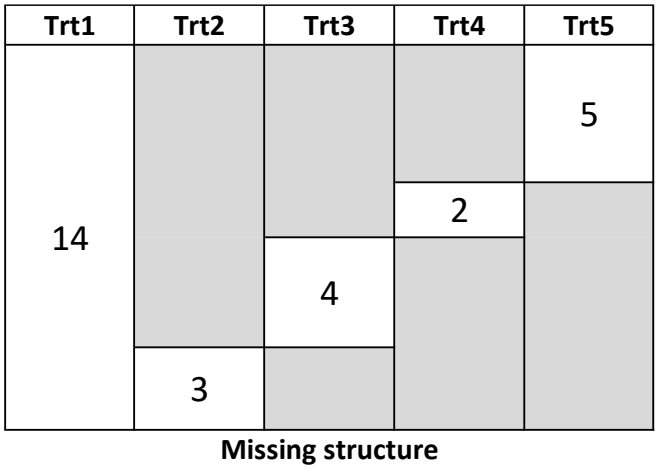


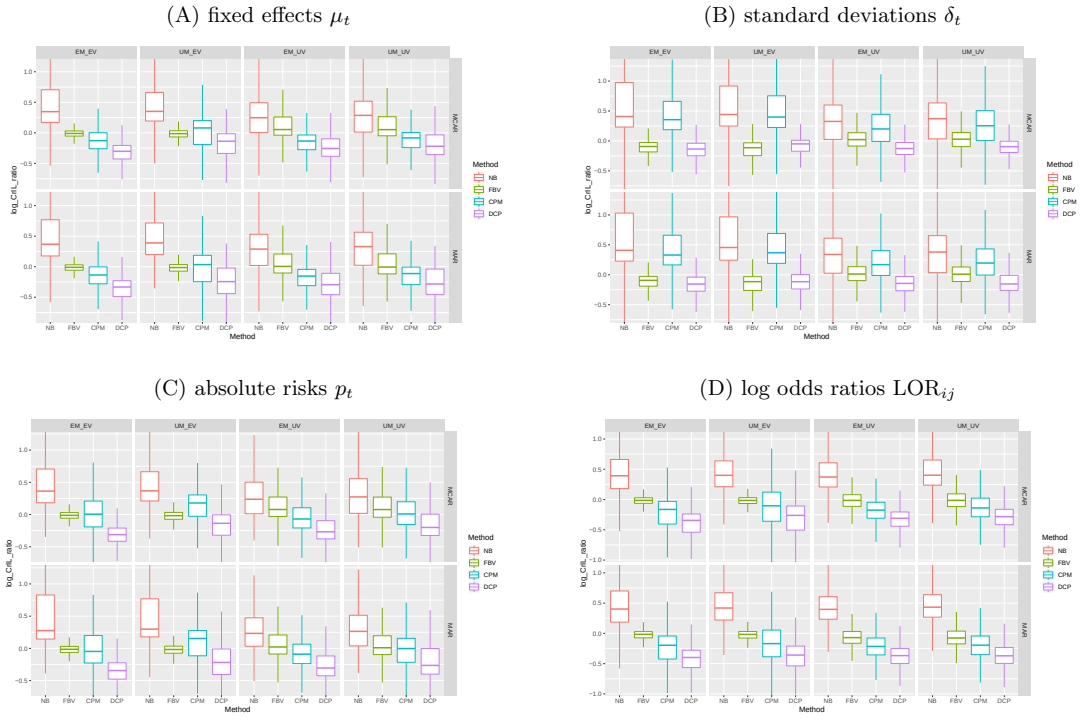
Table 4.4: Simulation results comparing data generated under four different scenarios (EM-EV, UM-EV, EM-UV, and UM-UV) with the MAR missingness of treatment arms. The bias and mean squared error of the posterior median and the coverage probability of the 95% credible interval are summarized for the five methods (NB, CPV, FBV, CPM, and DCP). For example, the value in the column of bias and the row of LOR_{ij} is calculated as $\sum_{i\neq j}|\mathrm{bias}(\mathrm{LOR}_{ij})|$; the value in the column of coverage probability and in the row of LOR_{ij} is calculated as $\sum_{i\neq j}(0.95-\mathrm{CP}(\mathrm{LOR}_{ij}))_+$.

Parameter	Truth	Bias					Mean squared error					Coverage probability				
		NB	CPV	FBV	CPM	DCP	NB	CPV	FBV	CPM	DCP	NB	CPV	FBV	CPM	DCP
Scenario EM-EV																
LOR_{ij}	.	1.82	0.35	0.48	0.60	0.25	4.84	3.85	3.83	2.06	2.05	0.00	0.00	0.00	0.00	0.00
μ_t	.	0.40	0.17	0.22	0.18	0.17	1.18	1.01	1.00	0.53	0.54	0.00	0.01	0.02	0.00	0.00
p_t	.	0.10	0.04	0.05	0.07	0.03	0.01	0.01	0.01	0.01	0.00	0.00	0.02	0.03	0.00	0.00
δ_t	.	1.83	0.53	0.69	1.22	0.38	2.17	0.82	0.87	1.37	0.55	0.07	0.05	0.11	0.02	0.05
LOR_{25}	−1.50	−0.41	0.03	−0.05	−0.13	0.01	0.68	0.35	0.38	0.25	0.23	1.00	0.98	0.98	1.00	0.99
p_5	0.19	0.03	−0.01	−0.00	0.01	−0.01	0.00	0.00	0.00	0.00	0.00	0.99	0.99	0.99	0.99	0.99
δ_5	0.30	0.32	−0.01	0.01	0.16	−0.04	0.21	0.03	0.03	0.08	0.02	0.93	0.99	0.98	0.96	0.99
Scenario UM-EV																
LOR_{ij}	.	1.81	0.36	0.46	2.21	2.11	4.83	3.82	3.80	2.63	2.60	0.00	0.00	0.00	0.00	0.03
μ_t	.	0.40	0.17	0.20	0.75	0.71	1.17	1.00	0.99	0.66	0.65	0.00	0.02	0.04	0.00	0.09
p_t	.	0.10	0.04	0.04	0.11	0.09	0.01	0.01	0.01	0.01	0.01	0.00	0.03	0.04	0.03	0.06
δ_t	.	1.83	0.45	0.63	1.13	0.25	2.17	0.76	0.76	1.28	0.51	0.07	0.07	0.18	0.00	0.07
LOR_{25}	−1.50	−0.41	0.03	−0.05	0.34	0.39	0.67	0.35	0.37	0.34	0.36	1.00	0.98	0.99	0.99	0.96
p_5	0.19	0.03	−0.00	−0.00	−0.01	−0.02	0.00	0.00	0.00	0.00	0.00	0.99	0.99	0.99	1.00	0.95
δ_5	0.30	0.32	0.01	0.03	0.17	−0.02	0.21	0.03	0.03	0.09	0.03	0.93	0.98	0.98	0.96	0.99
Scenario EM-UV																
LOR_{ij}	.	1.81	0.39	0.69	0.93	0.28	4.83	3.75	3.56	1.57	1.93	0.00	0.02	0.03	0.00	0.00
μ_t	.	0.40	0.11	0.24	0.23	0.08	1.17	0.97	0.92	0.39	0.50	0.00	0.02	0.04	0.00	0.00
p_t	.	0.10	0.04	0.11	0.08	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.02	0.10	0.00	0.00
δ_t	.	1.83	0.63	1.85	1.26	0.65	2.17	0.86	1.65	1.36	0.65	0.07	0.03	1.82	0.02	0.02
LOR_{25}	−1.50	−0.41	−0.03	−0.01	−0.21	−0.07	0.67	0.36	0.35	0.26	0.23	1.00	0.99	0.99	1.00	0.99
p_5	0.19	0.03	0.01	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.99	1.00	0.99	0.99	0.99
δ_5	0.30	0.32	0.23	0.48	0.22	0.19	0.21	0.12	0.37	0.12	0.09	0.93	0.94	0.55	0.95	0.95
Scenario UM-UV																
LOR_{ij}	.	1.82	0.46	0.94	2.03	1.83	4.84	3.64	3.47	1.74	2.23	0.00	0.03	0.06	0.00	0.03
μ_t	.	0.40	0.14	0.29	0.70	0.64	1.17	0.94	0.91	0.47	0.58	0.00	0.04	0.09	0.00	0.04
p_t	.	0.10	0.04	0.10	0.11	0.04	0.01	0.01	0.01	0.01	0.01	0.00	0.04	0.13	0.09	0.05
δ_t	.	1.83	0.70	1.83	1.10	0.77	2.17	0.82	1.61	1.27	0.61	0.07	0.02	1.73	0.01	0.03
LOR_{25}	−1.50	−0.41	0.02	0.08	0.26	0.33	0.68	0.34	0.32	0.23	0.28	1.00	0.99	0.99	1.00	0.99
p_5	0.19	0.03	0.01	0.03	0.00	−0.01	0.00	0.00	0.00	0.00	0.00	0.99	1.00	1.00	1.00	0.99
δ_5	0.30	0.32	0.23	0.49	0.20	0.17	0.21	0.13	0.42	0.11	0.08	0.93	0.95	0.58	0.95	0.96

Table 4.5: Simulation results comparing data generated under four different scenarios (EM-EV, UM-EV, EM-UV, and UM-UV) with the MCAR missingness of treatment arms. The bias and mean squared error of the posterior median and the coverage probability of the 95% credible interval are summarized for the five methods (NB, CPV, FBV, CPM, and DCP). For example, the value in the column of bias and in the row of LOR_{ij} is calculated as $\sum_{i\neq j}|\text{bias}(\text{LOR}_{ij})|$; the value in the column of coverage probability and in the row of LOR_{ij} is calculated as $\sum_{i\neq j}(0.95-\text{CP}(\text{LOR}_{ij}))_+$.

Parameter	Truth	Bias					Mean squared error					Coverage probability				
		NB	CPV	FBV	CPM	DCP	NB	CPV	FBV	CPM	DCP	NB	CPV	FBV	CPM	DCP
Scenario EM-EV																
LOR_{ij}	.	0.57	0.49	0.53	0.34	0.39	4.27	4.04	3.99	1.99	2.11	0.00	0.00	0.00	0.00	0.00
μ_t	.	0.23	0.17	0.19	0.15	0.16	1.09	1.04	1.02	0.50	0.54	0.00	0.01	0.02	0.00	0.00
p_t	.	0.08	0.03	0.04	0.07	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.02	0.02	0.00	0.00
δ_t	.	1.52	0.46	0.65	1.06	0.32	1.81	0.84	0.89	1.27	0.59	0.07	0.04	0.09	0.03	0.05
LOR_{25}	−1.50	−0.08	−0.05	−0.06	−0.06	−0.06	0.35	0.32	0.32	0.21	0.22	1.00	0.98	0.98	0.99	0.97
p_5	0.19	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.97	0.98	1.00	0.98
δ_5	0.30	0.26	−0.01	0.01	0.16	−0.04	0.16	0.03	0.03	0.09	0.03	0.93	0.98	0.98	0.95	0.99
Scenario UM-EV																
LOR_{ij}	.	0.57	0.48	0.53	2.02	1.38	4.27	3.98	3.92	2.51	2.33	0.00	0.00	0.00	0.00	0.02
μ_t	.	0.23	0.17	0.19	0.63	0.44	1.09	1.02	1.01	0.62	0.59	0.00	0.02	0.04	0.00	0.06
p_t	.	0.08	0.03	0.04	0.10	0.06	0.01	0.01	0.01	0.01	0.01	0.00	0.03	0.04	0.05	0.07
δ_t	.	1.52	0.39	0.58	1.23	0.25	1.81	0.75	0.76	1.40	0.54	0.07	0.07	0.16	0.05	0.07
LOR_{25}	−1.50	−0.08	−0.05	−0.06	0.25	0.17	0.35	0.31	0.31	0.27	0.24	1.00	0.98	0.98	0.98	0.96
p_5	0.19	0.01	0.00	0.00	−0.00	−0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.98	0.98	1.00	0.97
δ_5	0.30	0.26	0.00	0.03	0.24	0.02	0.16	0.03	0.03	0.13	0.03	0.93	0.97	0.97	0.93	0.97
Scenario EM-UV																
LOR_{ij}	.	0.57	0.31	0.24	0.47	0.38	4.26	3.78	3.64	1.35	1.96	0.00	0.05	0.09	0.00	0.00
μ_t	.	0.23	0.13	0.15	0.19	0.16	1.09	0.97	0.94	0.35	0.51	0.00	0.05	0.07	0.00	0.00
p_t	.	0.08	0.03	0.08	0.06	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.05	0.09	0.00	0.00
δ_t	.	1.52	0.65	1.87	1.07	0.69	1.81	0.85	1.70	1.24	0.67	0.07	0.03	1.87	0.04	0.02
LOR_{25}	−1.50	−0.08	−0.03	−0.01	−0.09	−0.08	0.35	0.30	0.30	0.19	0.20	1.00	0.99	0.99	0.99	0.99
p_5	0.19	0.01	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.99	0.99	0.99	0.98
δ_5	0.30	0.26	0.21	0.47	0.21	0.18	0.16	0.11	0.35	0.12	0.09	0.93	0.94	0.54	0.94	0.94
Scenario UM-UV																
LOR_{ij}	.	0.57	0.29	0.27	2.01	1.30	4.27	3.73	3.57	1.76	2.17	0.00	0.08	0.12	0.00	0.06
μ_t	.	0.23	0.11	0.14	0.64	0.43	1.09	0.96	0.92	0.45	0.56	0.00	0.07	0.10	0.00	0.06
p_t	.	0.08	0.03	0.08	0.11	0.03	0.01	0.01	0.01	0.01	0.01	0.00	0.07	0.11	0.09	0.05
δ_t	.	1.52	0.71	1.86	1.15	0.79	1.81	0.82	1.65	1.35	0.66	0.06	0.04	1.75	0.05	0.04
LOR_{25}	−1.50	−0.08	−0.01	0.01	0.22	0.17	0.35	0.29	0.28	0.21	0.21	1.00	0.98	0.98	0.99	0.97
p_5	0.19	0.01	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.99	0.99	1.00	0.99
δ_5	0.30	0.26	0.21	0.48	0.23	0.20	0.16	0.12	0.41	0.14	0.10	0.93	0.93	0.58	0.93	0.93

Figure 4.6: Simulation results comparing data generated under the four different scenarios (EM-EV, UM-EV, EM-UV, and UM-UV) with the two different missingness settings (MCAR and MAR). For each estimand, the log of the ratio of 95% credible interval length (CrIL) of each of the four methods (NB, FBV, CPM, and DCP) versus the CPV method are presented as box plots with whiskers showing the 1st and 99th percentiles (outliers not displayed). Subplots show results for: (a) fixed effects of log-odds μ_t; (b) standard deviations of log-odds δ_t; (c) absolute risks p_t; and (d) log odds ratios LOR_{ij}.



Chapter 5

Conclusion

In this chapter, we summarize the contributions of this thesis to analysis of the AB-NMA with binary outcomes in Section 5.1, and discuss future work in Section 5.2.

5.1 Summary of major findings

In this thesis, we have found two problems in analyzing arm-based Bayesian NMA with binary outcomes. First, the commonly-used conjugate prior for the covariance matrix, the inverse-Wishart (IW) distribution, generally leads to underestimation of correlations between treatment-specific log-odds, which are critical for borrowing strength across treatment arms to estimate treatment effects efficiently and to reduce potential bias. Second, because many NMAs collect only a few eligible clinical studies, the number of clinical studies involving each treatment is often small in an NMA, leading to unstable treatment-specific variance estimates in an AB-NMA 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.

In Chapter 2, we tried to solve the first problem by studying the IW prior's impact on AB-NMA results and comparing it with separation strategies (i.e., separate priors on variances and correlations) through simulation studies and case studies. For the separation strategy, different priors on the correlation matrix (e.g. restricted Wishart,

restricted inverse-Wishart, and the equal correlation prior) were also compared. Separation strategies, especially with the equal correlation prior, can improve estimation on correlations and treatment-specific log-odds. In Chapter 3, we proposed variance shrinkage methods (e.g., a hierarchical half-cauchy prior and a hierarchical inverse-gamma prior) to solve the second problem. We compared the variance shrinkage methods, homogeneous variance assumption, heterogeneous variance assumption, and the IW prior using extensive simulation studies, and found that only the hierarchical half-cauchy prior is robust under all simulation scenarios. Those results were further illustrated by a network meta-analysis on organised inpatient care for stroke. In Chapter 4, we proposed another approach to solve the second problem, incorporating single-arm studies into arm-based NMA. We performed comprehensive simulations to compare different borrowing strategies, especially fully borrowing and adaptively borrowing (i.e., fully borrowing on variances vs. commensurate prior on variances), and whether it is desirable to borrow both mean and variance information, i.e., comparing the double commensurate prior vs. commensurate prior on variances. We found that only the commensurate prior on variances was robust under all simulation scenarios. We performed a network meta-analysis on safety of immune checkpoint inhibitors to evaluate the performance of the proposed approach and to illustrate the importance of variance extrapolation when the number of clinical studies involving each treatment is relatively small.

Based on these simulation studies, case studies, and discussions, Figure 5.1 presents a road map for analyzing arm-based Bayesian NMA with binary outcomes. We recommend the separation strategy with the equal correlation prior rather than the IW prior for the AB-NMA. In the separation strategy, the traditional choices for the prior on variances would be either a homogeneous variance assumption or a heterogeneous variance assumption, and the literature offers no criteria for choosing or distinguishing between these two assumptions. We propose one criterion to determine choices for variance priors: whether any treatment $t \in \{1, \dots T\}$ in the NMA has less than 6 ($B_t \leq 5$) clinical studies involving this treatment. If not, then we could simply choose the heterogeneous variance assumption with a uniform prior $U(0, 5)$ for each treatment-specific standard deviation δ_t. On the other hand, if $B_t \leq 5$ we could either use hierarchical half-Cauchy prior proposed in Chapter 3 to shrink the variance in a data-dependent way if no single-arm studies are available, or we could use the commensurate prior on

variances proposed in Chapter 4 to bridge variance information between randomized controlled trials and single-arm studies.

5.2 Future work

Despite the contributions of this road map to the arm-based Bayesian NMA with binary outcomes, future work remains.

1. *Applying the variance shrinkage method in arm-based Bayesian NMA with other types of outcomes.* In Chapter 3, we discussed variance shrinkage method only for NMA with binary outcomes. However, 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 [65] is:

$$\begin{aligned} &\text{Level I: } \overline{y}_{kt} \sim N(\theta_{kt}, s_{kt}^2/n_{kt}), \quad t \in A_k,\ k = 1, \ldots, K; \\ &\text{Level II: } (\theta_{k1}, \ldots, \theta_{kT})' \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned} \tag{5.1}$$

   where θ_{kt} is the underlying mean outcome of the t^{th} treatment in the k^{th} study. Comparing Equations (3.1) and (5.1), the difference is on Level I (within study); because shrinkage is applied to $\delta_t,\ t = 1, \ldots, K$, 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.

2. *Applying the variance extrapolation method in arm-based Bayesian NMA with other types of outcomes.* Like the variance shrinkage method, various borrowing strategies discussed in Chapter 4, especially the commensurate prior on variances, could be applied to NMA with other types of outcomes.

3. *Applying the separation strategy with RIW and EQ priors on correlation matrix in contrast-based Bayesian NMA.* In CB-NMA, given a T-treatment NMA, the

observed dataset $\{D_1, \ldots, D_K\}$ with $D_k = \{(r_{kt}, n_{kt}); t \in A_k\}$ can be generated by the following specification [5]:

$$\begin{aligned} r_{kt} &\sim \text{Binomial}(n_{kt}, p_{kt}), \quad t \in A_k, \quad k = 1, \ldots, K; \\ \text{logit}(p_{k2}) = \alpha_b &\sim N(\mu_\alpha, \sigma_\alpha^2); \\ \text{logit}(p_{kj}) - \text{logit}(p_{k2}) = \delta_{k2j} &\sim N(d_{2j}, \varsigma_{2j}^2), \quad j = 1, 3, 4, \ldots, T; \\ \text{corr}(\delta_{k2i}, \delta_{k2j}) &= \gamma_{ij}^{(2)}, \quad i \neq j \neq 2, \end{aligned} \tag{5.2}$$

where treatment 2 is assumed to be the baseline for all studies for simplicity. The baseline treatment effect α_b is generally treated as a nuisance parameter, though here we assume it follows a normal distribution $N(\mu_\alpha, \sigma_\alpha^2)$ to help readers understand the connection between AB-NMA and CB-NMA. The random effect δ_{k2j} is the study-specific relative effect of treatment j vs. the baseline treatment 2, which follows a normal distribution with overall mean effect d_{2j} and between-trial variance ς_{2j}^2. Also, $\gamma_{ij}^{(2)}$ denotes the correlation between the two contrasts i vs. 2 and j vs. 2 in the k^{th} study. Note that for all treatment pairs (m, n), the random effects $\{\delta_{kmn}, k = 1, \ldots, K\}$ are conditionally independent given the true d_{mn} and ς_{mn}^2; this is the exchangeability assumption for the relative effects. Obviously, this model can be written as one form of the AB model in Equation (2.1) with mean $(\mu_1, \mu_2, \mu_3, \ldots, \mu_T) = (\mu_\alpha + d_{21}, \mu_\alpha, \mu_\alpha + d_{23}, \ldots, \mu_\alpha + d_{2T})$, variances $(\delta_1^2, \delta_2^2, \delta_3^2, \ldots, \delta_T^2) = (\sigma_\alpha^2 + \varsigma_{21}^2, \sigma_\alpha^2, \sigma_\alpha^2 + \varsigma_{23}^2, \ldots, \sigma_\alpha^2 + \varsigma_{2T}^2)$, and correlation matrix $\mathbf{P}$ with entries $(\rho_{2i}, \rho_{ij}) = \left(\frac{\sigma_\alpha}{\sqrt{\sigma_\alpha^2 + \varsigma_{2i}^2}}, \frac{\sigma_\alpha^2 + \gamma_{ij}^{(2)} \varsigma_{2i} \varsigma_{2j}}{\sqrt{(\sigma_\alpha^2 + \varsigma_{2i}^2)(\sigma_\alpha^2 + \varsigma_{2j}^2)}} \right)$ for $i \neq j \neq 2$. As we can see, the $T \times T$ covariance matrix in the AB model is closely related to the CB model's $(T-1) \times (T-1)$ covariance matrix, so we can probably apply corresponding RIW and EQ priors in the CB approach. In Lu and Ades' paper [6], they discussed the separation strategy for CB-NMA's $(T-1) \times (T-1)$ covariance matrix. However, they only proposed spherical parameterization methods for the correlation matrix, and as we discussed in Section 2.6 the marginal distributions of the ρ_{ij} depend strongly on the indexes i and j in this parameterization approach.

4. *Applying the variance shrinkage and the variance extrapolation methods in CB-NMA.* As we discussed in Chapter 1, and as we can see from Equation (5.2), if we have too few comparisons between treatments 2 and i (i.e., less than 5, which

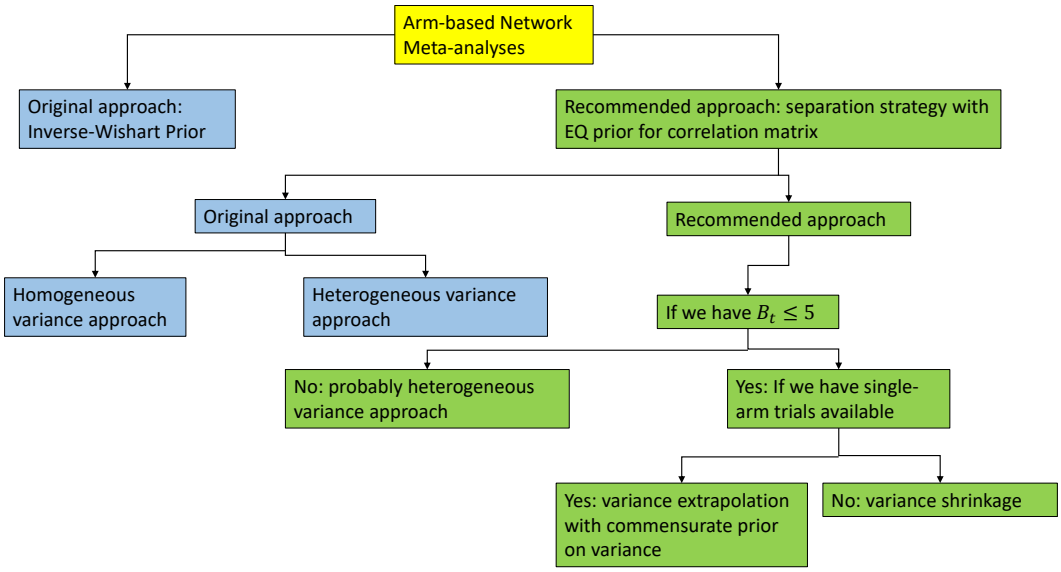
is quite prevalent in NMA), then as with AB-NMA, the posterior distribution of ς_{2i} will certainly be dominated by prior information. We could apply the idea of variance shrinkage and variance extrapolation methods to CB-NMA. However, as discussed in Lu and Ades [6], the triangle inequalities on contrast standard deviations

$$|\varsigma_{2i} - \varsigma_{2j}| \leq \varsigma_{ij} \leq \varsigma_{2i} + \varsigma_{2j},\ i \neq j \neq 2$$

could complicate prior specifications.

5. *Information borrowing strategies on meta-regression and network meta-regression.* We discussed the different information borrowing strategies in AB-NMA by comparing fully borrowing and adaptively borrowing, and determing whether to borrow both mean and variance information. On top of the models considered in ordinary meta-analyses, meta-regression and network meta-regression consider extra covariates. We could also discuss potential strategies for information extrapolation in these models, especially for covariates and covariates adjusted effects.

Figure 5.1: Road map for arm-based Bayesian NMA with binary outcomes



References

[1] Nikolakopoulou, A., Chaimani, A., Veroniki, A. A., et al. Characteristics of networks of interventions: A description of a database of 186 published networks. *PLoS ONE*, 9(1):e86754, 2014.

[2] Trinder, L. A critical appraisal of evidence-based practice. In *Evidence-based Practice*, pages 212–241. Blackwell Science Ltd, 2000.

[3] Egger, M., Davey Smith, G., and Altman, D. G., editors. *Systematic Reviews in Health Care*. BMJ Publishing Group, 2001.

[4] Mills, E. J., Ioannidis, J. P. A., Thorlund, K., et al. How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA*, 308(12):1246–1253, 2012.

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

[6] Lu, G. and Ades, A. E. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics*, 10(4):792–805, 2009.

[7] Zhang, J., Carlin, B. P., Neaton, J. D., et al. Network meta-analysis of randomized clinical trials: reporting the proper summaries. *Clinical Trials*, 11(2):246–262, 2014.

[8] Bucher, H. C., Guyatt, G. H., Griffith, L. E., and Walter, S. D. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology*, 50(6):683–691, 1997.

[9] Rücker, G. and Schwarzer, G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology*, 15(1):58, 2015.

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

[11] Hong, H., Chu, H., Zhang, J., and Carlin, B. P. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods*, 7(1):6–22, 2016.

[12] Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., and Shen, W. Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of computational and graphical statistics*, 26:121–133, 2017.

[13] Dias, S. and Ades, A. E. Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods*, 7(1):23–28, 2016.

[14] Hong, H., Chu, H., Zhang, J., and Carlin, B. P. Rejoinder to the discussion of "a Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons", by S. Dias and A.E. Ades. *Research Synthesis Methods*, 7(1):29–33, 2016.

[15] Béliveau, A., Goring, S., Platt, R. W., and Gustafson, P. Network meta-analysis of disconnected networks: how dangerous are random baseline treatment effects? *Research Synthesis Methods*, 8(4):465–474, 2017.

[16] Lin, L., Chu, H., and Hodges, J. S. Sensitivity to excluding treatments in network meta-analysis. *Epidemiology*, 27(4):562–569, 2016.

[17] Barnard, J., McCulloch, R., and Meng, X. L. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1311, 2000.

[18] Caldwell, D. M., Ades, A. E., and Higgins, J. P. T. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*, 331(7521):897–900, 2005.

[19] Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534, 2006.

[20] Alvarez, I., Niemi, J., and Simpson, M. Bayesian inference for a covariance matrix. *Conference on Applied Statistics in Agriculture*, 26(1):71–82, 2014.

[21] Kontopantelis, E., Springate, D. A., and Reeves, D. A re-analysis of the cochrane library data: The dangers of unobserved heterogeneity in meta-analyses. *PLoS ONE*, 8(7):e69930, 2013.

[22] Mathes, T. and Kuss, O. A comparison of methods for meta-analysis of a small number of studies with binary outcomes. *Research Synthesis Methods*, 2018.

[23] Dias, S., Sutton, A. J., Ades, A. E., and Welton, N. J. 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(5):607–617, 2013.

[24] James, W. and Stein, C. Estimation with quadratic loss. In *Springer Series in Statistics*, pages 443–460. Springer New York, 1992.

[25] Hwang, J. T. G., Qiu, J., and Zhao, Z. Empirical Bayes confidence intervals shrinking both means and variances. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):265–285, 2009.

[26] Hwang, J. G. and Liu, P. Optimal tests shrinking both means and variances applicable to microarray data analysis. *Statistical Applications in Genetics and Molecular Biology*, 9(1), 2010.

[27] Zhao, Z. Double shrinkage empirical Bayesian estimation for unknown and unequal variances. *Statistics and Its Interface*, 3(4):533–541, 2010.

[28] Hueber, W., Sands, B. E., Lewitzky, S., et al. Secukinumab, a human anti-IL-17a monoclonal antibody, for moderate to severe crohn's disease: unexpected results

of a randomised, double-blind placebo-controlled trial. *Gut*, 61(12):1693–1700, 2012.

[29] Chen, M.-H., Ibrahim, J. G., Lam, P., Yu, A., and Zhang, Y. Bayesian design of noninferiority trials for medical devices using historical data. *Biometrics*, 67(3):1163–1170, 2011.

[30] Gamalo, M. A., Tiwari, R. C., and LaVange, L. M. Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products. *Pharmaceutical Statistics*, 13(1):25–40, 2013.

[31] Gamalo-Siebers, M., Savic, J., Basu, C., et al. Statistical modeling for bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharmaceutical Statistics*, 16(4):232–249, 2017.

[32] Begg, C. B. and Pilote, L. A model for incorporating historical controls into a meta-analysis. *Biometrics*, 47:899–906, 1991.

[33] Li, Z. and Begg, C. B. Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. *Journal of the American Statistical Association*, 89(428), 1994.

[34] Zhang, J., Ko, C.-W., Nie, L., Chen, Y., and Tiwari, R. Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. *Statistical Methods in Medical Research*, 28(5):1293–1310, 2018.

[35] Röver, C., Wandel, S., and Friede, T. Model averaging for robust extrapolation in evidence synthesis. *Statistics in Medicine*, 38(4):674–694, 2018.

[36] Efthimiou, O., Mavridis, D., Debray, T. P. A., et al. Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine*, 36(8):1210–1226, 2017.

[37] Turner, R. M., Domínguez-Islas, C. P., Jackson, D., Rhodes, K. M., and White, I. R. Incorporating external evidence on between-trial heterogeneity in network meta-analysis. *Statistics in Medicine*, 38(8):1321–1335, 2019.

[38] Leahy, J., Thom, H., Jansen, J. P., et al. Incorporating single-arm evidence into a network meta-analysis using aggregate level matching: Assessing the impact. *Statistics in Medicine*, 38(14):2505–2523, 2019.

[39] Hong, H., Fu, H., and Carlin, B. P. Power and commensurate priors for synthesizing aggregate and individual patient level data in network meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(4):1047–1069, 2018.

[40] White, I. R., Turner, R. M., Karahalios, A., and Salanti, G. A comparison of arm-based and contrast-based models for network meta-analysis. *Statistics in Medicine*, 38(27):5197–5213, 2019.

[41] O'Malley, A. J. and Zaslavsky, A. M. Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association*, 103(484):1405–1418, 2008.

[42] Chen, Z. and Dunson, D. B. Random effects selection in linear mixed models. *Biometrics*, 59(4):762–769, 2003.

[43] Lewandowski, D., Kurowicka, D., and Joe, H. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001, 2009.

[44] Zeger, S. L., Liang, K.-Y., and Albert, P. S. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44(4):1049–1060, 1988.

[45] Agresti, A. *Categorical data analysis*. John Wiley & Sons, Hoboken, NJ, 3 edition, 2013.

[46] Huang, A. and Wand, M. P. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452, 2013.

[47] Lin, L., Zhang, J., Hodges, J. S., and Chu, H. Performing arm-based network meta-analysis in r with the pcnetmeta package. *Journal of Statistical Software*, 80(5):1–25, 2017.

[48] Carpenter, B., Gelman, A., Hoffman, M. D., et al. Stan: a probabilistic programming language. *Journal of Statistical Software*, 76(1):1–29, 2017.

[49] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[50] Mills, E. J., Wu, P., Spurden, D., Ebbert, J. O., and Wilson, K. Efficacy of pharmacotherapies for short-term smoking abstinance: a systematic review and meta-analysis. *Harm Reduction Journal*, 6(1):25, 2009.

[51] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[52] Thijs, V., Lemmens, R., and Fieuws, S. Network meta-analysis: simultaneous meta-analysis of common antiplatelet regimens after transient ischaemic attack or stroke. *European Heart Journal*, 29(9):1086–1092, 2008.

[53] Hofmeyr, G. J., Gülmezoglu, A. M., Novikova, N., et al. Misoprostol to prevent and treat postpartum haemorrhage: a systematic review and meta-analysis of maternal deaths and dose-related effects. *Bulletin of the World Health Organization*, 87(9):666–667, 2009.

[54] Pinheiro, J. C. and Bates, D. M. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296, 1996.

[55] Wei, Y. H. and Higgins, J. P. T. Bayesian multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*, 32(17):2911–2934, 2013.

[56] Lin, L. and Chu, H. Bayesian multivariate meta-analysis of multiple factors. *Research Synthesis Methods*, 9(2):261–272, 2018.

[57] Senn, S. Trying to be precise about vagueness. *Statistics in Medicine*, 26(7):1417–1430, 2007.

[58] Houwelingen, H. C. V., Zwinderman, K. H., and Stijnen, T. A bivariate approach to meta-analysis. *Statistics in Medicine*, 12(24):2273–2284, 1993.

[59] Senn, S. Hans van houwelingen and the art of summing up. *Biometrical Journal*, 52(1):85–94, 2010.

[60] Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., and Higgins, J. P. T. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology*, 41(3):818–827, 2012.

[61] 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. *The Lancet*, 380(9859):2095–2128, 2012.

[62] Murray, C. J. L., 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. *The Lancet*, 380(9859):2197–2223, 2012.

[63] Martin, P. Stroke units: An evidence based approach. *Journal of Neurology, Neurosurgery & Psychiatry*, 66(3):412–412, 1999.

[64] Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. *Cochrane Database of Systematic Reviews*, (4), 2007.

[65] Zhang, J., Fu, H., and Carlin, B. P. Detecting outlying trials in network meta-analysis. *Statistics in Medicine*, 34(19):2695–2707, 2015.

[66] Salanti, G., Ades, A., and Ioannidis, J. P. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology*, 64(2):163–171, 2011.

[67] Watanabe, S. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.*, 11:3571–3594, 2010.

[68] Lunn, D., Jackson, C., Best, N., Spiegelhalter, D., and Thomas, A. *The BUGS Book*. Taylor & Francis Inc, 2010.

[69] Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360–1383, 2008.

[70] Ghosh, J., Li, Y., and Mitra, R. On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, 13(2):359–383, 2018.

[71] Chen, M.-H. and Ibrahim, J. G. Power prior distributions for regression models. *Statistical Science*, 15(1):46–60, 2000.

[72] Duan, Y., Ye, K., and Smith, E. P. Evaluating water quality using power priors to incorporate historical information. *Environmetrics*, 17(1):95–106, 2005.

[73] Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. The power prior: theory and applications. *Statistics in Medicine*, 34(28):3724–3749, 2015.

[74] Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3):1047–1056, 2011.

[75] Hobbs, B. P., Sargent, D. J., and Carlin, B. P. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, 7(3):639–674, 2012.

[76] Murray, T. A., Hobbs, B. P., and Carlin, B. P. Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. *The Annals of Applied Statistics*, 9(3):1549–1570, 2015.

[77] Schmidli, H., Gsteiger, S., Roychoudhury, S., et al. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014.

[78] Kaizer, A. M., Koopmeiners, J. S., and Hobbs, B. P. Bayesian hierarchical modeling based on multisource exchangeability. *Biostatistics*, 19(2):169–184, 2017.

[79] Kaizer, A. M., Hobbs, B. P., and Koopmeiners, J. S. A multi-source adaptive platform design for testing sequential combinatorial therapeutic strategies. *Biometrics*, 74(3):1082–1094, 2018.

[80] Johnson, D. B., Chandra, S., and Sosman, J. A. Immune checkpoint inhibitor toxicity in 2018. *JAMA*, 320(16):1702, 2018.

[81] Hargadon, K. M., Johnson, C. E., and Williams, C. J. Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors. *International Immunopharmacology*, 62:29–39, 2018.

[82] Xu, C., Chen, Y.-P., Du, X.-J., et al. Comparative safety of immune checkpoint inhibitors in cancer: systematic review and network meta-analysis. *BMJ*, page k4226, 2018.

[83] Ribas, A., Kefford, R., Marshall, M. A., et al. Phase III randomized clinical trial comparing tremelimumab with standard-of-care chemotherapy in patients with advanced melanoma. *Journal of Clinical Oncology*, 31(5):616–622, 2013.

[84] Motzer, R. J., Rini, B. I., McDermott, D. F., et al. Nivolumab for metastatic renal cell carcinoma: Results of a randomized phase II trial. *Journal of Clinical Oncology*, 33(13):1430–1437, 2015.

[85] Bang, Y.-J., Cho, J. Y., Kim, Y. H., et al. Efficacy of sequential ipilimumab monotherapy versus best supportive care for unresectable locally advanced/metastatic gastric or gastroesophageal junction cancer. *Clinical Cancer Research*, 23(19):5671–5678, 2017.

[86] Hersh, E. M., O'Day, S. J., Powderly, J., et al. A phase II multicenter study of ipilimumab with or without dacarbazine in chemotherapy-naïve patients with advanced melanoma. *Investigational New Drugs*, 29(3):489–498, 2010.

[87] de Valpine, P., Turek, D., Paciorek, C. J., et al. Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2):403–413, 2017.

[88] Geisser, S. and Eddy, W. F. A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153–160, 1979.

[89] Hanson, T. E., Branscum, A. J., and Johnson, W. O. Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime Data Analysis*, 17(1):3–28, 2011.

[90] Weber, J., Mandala, M., Vecchio, M. D., et al. Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. *New England Journal of Medicine*, 377(19):1824–1835, 2017.

[91] Larkin, J., Chiarion-Sileni, V., Gonzalez, R., et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *New England Journal of Medicine*, 373(1):23–34, 2015.

[92] Wolchok, J. D., Chiarion-Sileni, V., Gonzalez, R., et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *New England Journal of Medicine*, 377(14):1345–1356, 2017.

[93] Robert, C., Schachter, J., Long, G. V., et al. Pembrolizumab versus ipilimumab in advanced melanoma. *New England Journal of Medicine*, 372(26):2521–2532, 2015.

[94] Schachter, J., Ribas, A., Long, G. V., et al. Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006). *The Lancet*, 390(10105):1853–1862, 2017.

[95] Rittmeyer, A., Barlesi, F., Waterkamp, D., et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet*, 389(10066):255–265, 2017.

[96] Weber, J. S., D'Angelo, S. P., Minor, D., et al. Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *The Lancet Oncology*, 16(4):375–384, 2015.

[97] Larkin, J., Minor, D., D'Angelo, S., et al. Overall survival in patients with advanced melanoma who received nivolumab versus investigator's choice chemotherapy in CheckMate 037: A randomized, controlled, open-label phase III trial. *Journal of Clinical Oncology*, 36(4):383–390, 2018.

[98] Robert, C., Thomas, L., Bondarenko, I., et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *New England Journal of Medicine*, 364(26):2517–2526, 2011.

[99] Maio, M., Grob, J.-J., Aamdal, S., et al. Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *Journal of Clinical Oncology*, 33(10):1191–1196, 2015.

[100] Ferris, R. L., Blumenschein, G., Fayette, J., et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *New England Journal of Medicine*, 375(19):1856–1867, 2016.

[101] Govindan, R., Szczesna, A., Ahn, M.-J., et al. Phase III trial of ipilimumab combined with paclitaxel and carboplatin in advanced squamous non–small-cell lung cancer. *Journal of Clinical Oncology*, 35(30):3449–3457, 2017.

[102] Motzer, R. J., Escudier, B., McDermott, D. F., et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *New England Journal of Medicine*, 373(19):1803–1813, 2015.

[103] Carbone, D. P., Reck, M., Paz-Ares, L., et al. First-line nivolumab in stage IV or recurrent non–small-cell lung cancer. *New England Journal of Medicine*, 376(25):2415–2426, 2017.

[104] Bellmunt, J., de Wit, R., Vaughn, D. J., et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *New England Journal of Medicine*, 376(11):1015–1026, 2017.

[105] Ascierto, P. A., Vecchio, M. D., Robert, C., et al. Ipilimumab 10 mg/kg versus ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma: a randomised, double-blind, multicentre, phase 3 trial. *The Lancet Oncology*, 18(5):611–622, 2017.

[106] Reck, M., Rodríguez-Abreu, D., Robinson, A. G., et al. Pembrolizumab versus chemotherapy for PD-l1–positive non–small-cell lung cancer. *New England Journal of Medicine*, 375(19):1823–1833, 2016.

[107] Reck, M., Luft, A., Szczesna, A., et al. Phase III randomized trial of ipilimumab plus etoposide and platinum versus placebo plus etoposide and platinum in extensive-stage small-cell lung cancer. *Journal of Clinical Oncology*, 34(31):3740–3748, 2016.

[108] Langer, C. J., Gadgeel, S. M., Borghaei, H., et al. Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. *The Lancet Oncology*, 17(11):1497–1508, 2016.

[109] Postow, M. A., Chesney, J., Pavlick, A. C., et al. Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. *New England Journal of Medicine*, 372(21):2006–2017, 2015.

[110] Hodi, F. S., Chesney, J., Pavlick, A. C., et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *The Lancet Oncology*, 17(11):1558–1568, 2016.

[111] Herbst, R. S., Baas, P., Kim, D.-W., et al. Pembrolizumab versus docetaxel for previously treated, PD-l1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *The Lancet*, 387(10027):1540–1550, 2016.

[112] Fehrenbacher, L., Spira, A., Ballinger, M., et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *The Lancet*, 387(10030):1837–1846, 2016.

[113] Robert, C., Long, G. V., Brady, B., et al. Nivolumab in previously untreated melanoma without BRAF mutation. *New England Journal of Medicine*, 372(4):320–330, 2015.

[114] Ribas, A., Puzanov, I., Dummer, R., et al. Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): a randomised, controlled, phase 2 trial. *The Lancet Oncology*, 16(8):908–918, 2015.

[115] Brahmer, J., Reckamp, K. L., Baas, P., et al. Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. *New England Journal of Medicine*, 373(2):123–135, 2015.

[116] Borghaei, H., Paz-Ares, L., Horn, L., et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. *New England Journal of Medicine*, 373(17):1627–1639, 2015.

[117] Powles, T., Durán, I., van der Heijden, M. S., et al. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *The Lancet*, 391(10122):748–757, 2018.

[118] Reck, M., Bondarenko, I., Luft, A., et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line therapy in extensive-disease-small-cell lung cancer: results from a randomized, double-blind, multicenter phase 2 trial†. *Annals of Oncology*, 24(1):75–83, 2012.

[119] Lynch, T. J., Bondarenko, I., Luft, A., et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non–small-cell lung cancer: Results from a randomized, double-blind, multicenter phase II study. *Journal of Clinical Oncology*, 30(17):2046–2054, 2012.

[120] Hamid, O., Schmidt, H., Nissan, A., et al. A prospective phase II trial exploring the association between tumor microenvironment biomarkers and clinical activity of ipilimumab in advanced melanoma. *Journal of Translational Medicine*, 9(1), 2011.

[121] Wolchok, J. D., Neyns, B., Linette, G., et al. Ipilimumab monotherapy in patients with pretreated advanced melanoma: a randomised, double-blind, multicentre, phase 2, dose-ranging study. *The Lancet Oncology*, 11(2):155–164, 2010.

[122] Giaccone, G., Kim, C., Thompson, J., et al. Pembrolizumab in patients with thymic carcinoma: a single-arm, single-centre, phase 2 study. *The Lancet Oncology*, 19(3):347–355, 2018.

[123] Goldberg, S. B., Gettinger, S. N., Mahajan, A., et al. Pembrolizumab for patients with melanoma or non-small-cell lung cancer and untreated brain metastases: early analysis of a non-randomised, open-label, phase 2 trial. *The Lancet Oncology*, 17(7):976–983, 2016.

[124] Haag, G., Zoernig, I., Hassel, J., et al. Phase II trial of ipilimumab in melanoma patients with preexisting humoural immune response to NY-ESO-1. *European Journal of Cancer*, 90:122–129, 2018.

[125] Nghiem, P. T., Bhatia, S., Lipson, E. J., et al. PD-1 blockade with pembrolizumab in advanced merkel-cell carcinoma. *New England Journal of Medicine*, 374(26):2542–2552, 2016.

[126] Younes, A., Santoro, A., Shipp, M., et al. Nivolumab for classical hodgkin's lymphoma after failure of both autologous stem-cell transplantation and brentuximab vedotin: a multicentre, multicohort, single-arm phase 2 trial. *The Lancet Oncology*, 17(9):1283–1294, 2016.

[127] Amin, A., Lawson, D. H., Salama, A. K., et al. Phase II study of vemurafenib followed by ipilimumab in patients with previously untreated BRAF-mutated metastatic melanoma. *Journal for ImmunoTherapy of Cancer*, 4(44), 2016.

[128] Giacomo, A. M. D., Ascierto, P. A., Pilla, L., et al. Ipilimumab and fotemustine in patients with advanced melanoma (NIBIT-m1): an open-label, single-arm phase 2 trial. *The Lancet Oncology*, 13(9):879–886, 2012.

[129] Hamanishi, J., Mandai, M., Ikeda, T., et al. Safety and antitumor activity of anti–PD-1 antibody, nivolumab, in patients with platinum-resistant ovarian cancer. *Journal of Clinical Oncology*, 33(34):4015–4022, 2015.

[130] Hida, T., Nishio, M., Nogami, N., et al. Efficacy and safety of nivolumab in japanese patients with advanced or recurrent squamous non-small cell lung cancer. *Cancer Science*, 108(5):1000–1006, 2017.

[131] Kudo, T., Hamamoto, Y., Kato, K., et al. Nivolumab treatment for oesophageal squamous-cell carcinoma: an open-label, multicentre, phase 2 trial. *The Lancet Oncology*, 18(5):631–639, 2017.

[132] Maruyama, D., Hatake, K., Kinoshita, T., et al. Multicenter phase II study of nivolumab in japanese patients with relapsed or refractory classical hodgkin lymphoma. *Cancer Science*, 108(5):1007–1012, 2017.

[133] Nishio, M., Hida, T., Atagi, S., et al. Multicentre phase II study of nivolumab in japanese patients with advanced or recurrent non-squamous non-small cell lung cancer. *ESMO Open*, 2(Suppl 1):e000108, 2017.

[134] Overman, M. J., Lonardi, S., Wong, K. Y. M., et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair–deficient/microsatellite instability–high metastatic colorectal cancer. *Journal of Clinical Oncology*, 36(8):773–779, 2018.

[135] Sharma, P., Retz, M., Siefker-Radtke, A., et al. Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *The Lancet Oncology*, 18(3):312–322, 2017.

[136] Yamazaki, N., Kiyohara, Y., Uhara, H., et al. Phase II study of ipilimumab monotherapy in japanese patients with advanced melanoma. *Cancer Chemotherapy and Pharmacology*, 76(5):997–1004, 2015.

[137] Yamazaki, N., Kiyohara, Y., Uhara, H., et al. Efficacy and safety of nivolumab in japanese patients with previously untreated advanced melanoma: A phase II study. *Cancer Science*, 108(6):1223–1230, 2017.

[138] Yamazaki, N., Uhara, H., Fukushima, S., et al. Phase II study of the immune-checkpoint inhibitor ipilimumab plus dacarbazine in japanese patients with previously untreated, unresectable or metastatic melanoma. *Cancer Chemotherapy and Pharmacology*, 76(5):969–975, 2015.

[139] Zimmer, L., Vaubel, J., Mohr, P., et al. Phase II DeCOG-study of ipilimumab in pretreated and treatment-naïve patients with metastatic uveal melanoma. *PLOS ONE*, 10(3):e0118564, 2015.

[140] Zimmer, L., Eigentler, T. K., Kiecker, F., et al. Open-label, multicenter, single-arm phase II DeCOG-study of ipilimumab in pretreated patients with different subtypes of metastatic melanoma. *Journal of Translational Medicine*, 13(1), 2015.

[141] Balar, A. V., Galsky, M. D., Rosenberg, J. E., et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *The Lancet*, 389(10064):67–76, 2017.

[142] Bauml, J., Seiwert, T. Y., Pfister, D. G., et al. Pembrolizumab for platinum- and cetuximab-refractory head and neck cancer: Results from a single-arm, phase II study. *Journal of Clinical Oncology*, 35(14):1542–1549, 2017.

[143] O'Day, S. J., Maio, M., Chiarion-Sileni, V., et al. Efficacy and safety of ipilimumab monotherapy in patients with pretreated advanced melanoma: a multicenter single-arm phase II study. *Annals of Oncology*, 21(8):1712–1717, 2010.

[144] Peters, S., Gettinger, S., Johnson, M. L., et al. Phase II trial of atezolizumab as first-line or subsequent therapy for patients with programmed death-ligand 1–selected advanced non–small-cell lung cancer (BIRCH). *Journal of Clinical Oncology*, 35(24):2781–2789, 2017.

[145] Rizvi, N. A., Mazières, J., Planchard, D., et al. Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *The Lancet Oncology*, 16(3):257–265, 2015.

[146] Rosenberg, J. E., Hoffman-Censits, J., Powles, T., et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet*, 387(10031):1909–1920, 2016.

[147] Kang, Y.-K., Boku, N., Satoh, T., et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant

of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet*, 390(10111):2461–2471, 2017.

[148] Kwon, E. D., Drake, C. G., Scher, H. I., et al. Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised, double-blind, phase 3 trial. *The Lancet Oncology*, 15(7):700–712, 2014.

[149] James, A. T. Distributions of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, 35(2):475–501, 1964.

[150] Abramowitz, M. and Stegun, I. A., editors. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables (Dover Books on Mathematics)*. Dover Publications, 1965.

[151] Eaton, M. L. *Multivariate Statistics: A Vector Space Approach (Wiley Series in Probability and Statistics)*. John Wiley & Sons Inc, 1983.

Appendix A

Supplementary Materials

A.1 Diagnostic plots for applying the HHC method in the case study

A.1.1 Absolute risks p_t

Figure A.1: Trace plots of absolute risks p_t by chain

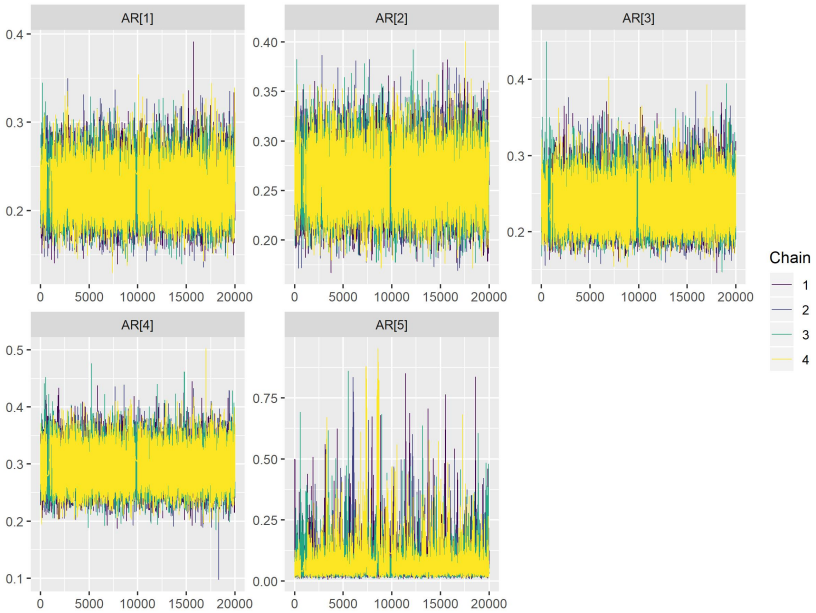
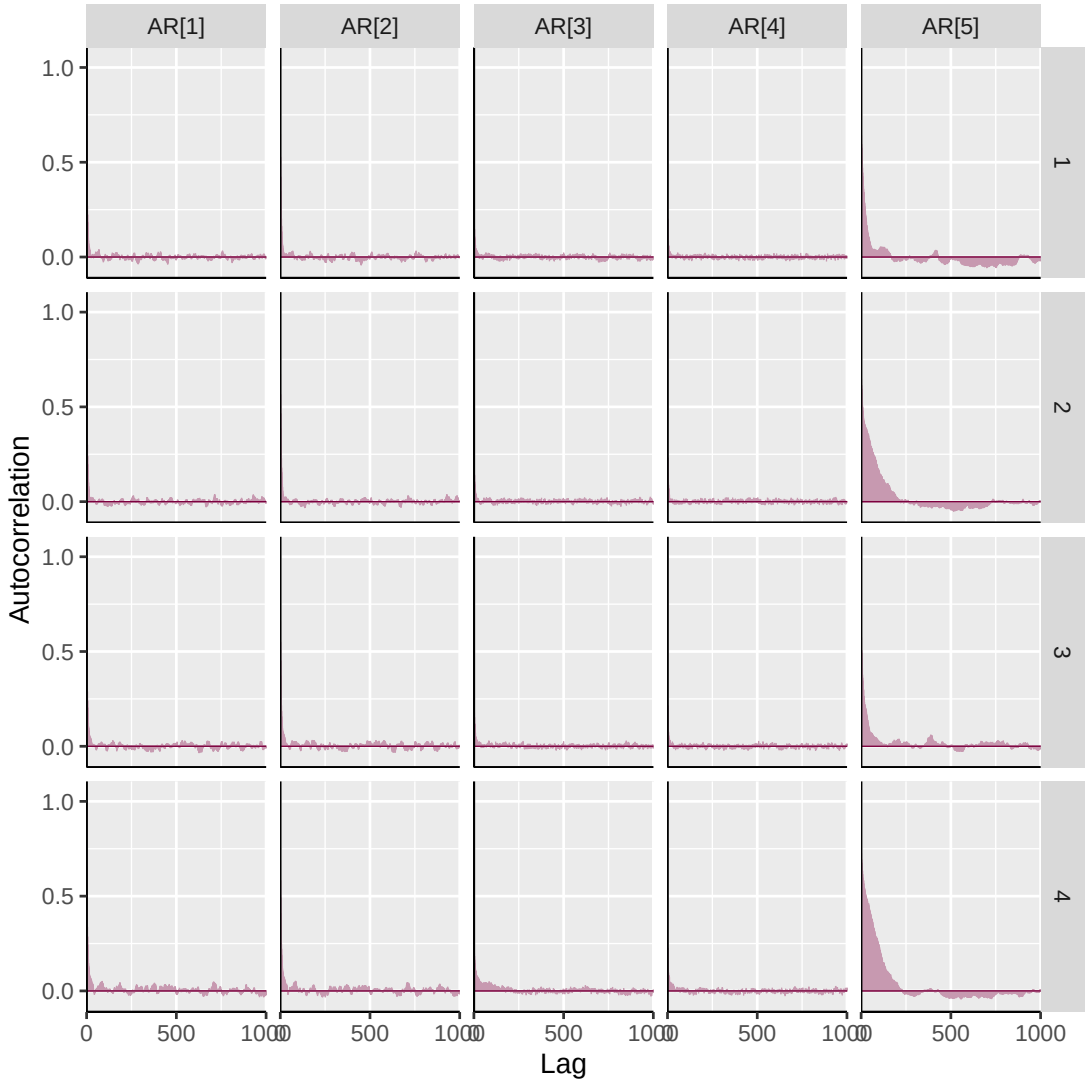


Figure A.2: Autocorrelation plots of absolute risks p_t by chain



A.1.2 Standard deviations δ_t

Figure A.3: Trace plots of standard deviations δ_t by chain

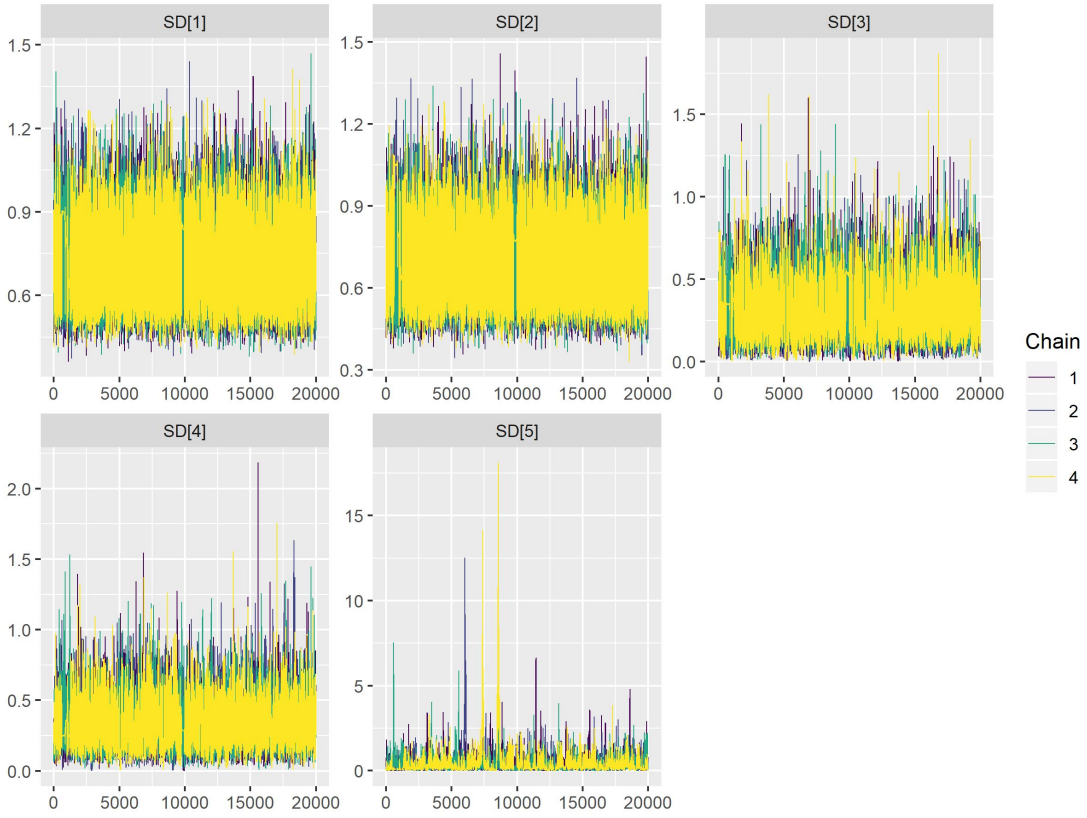
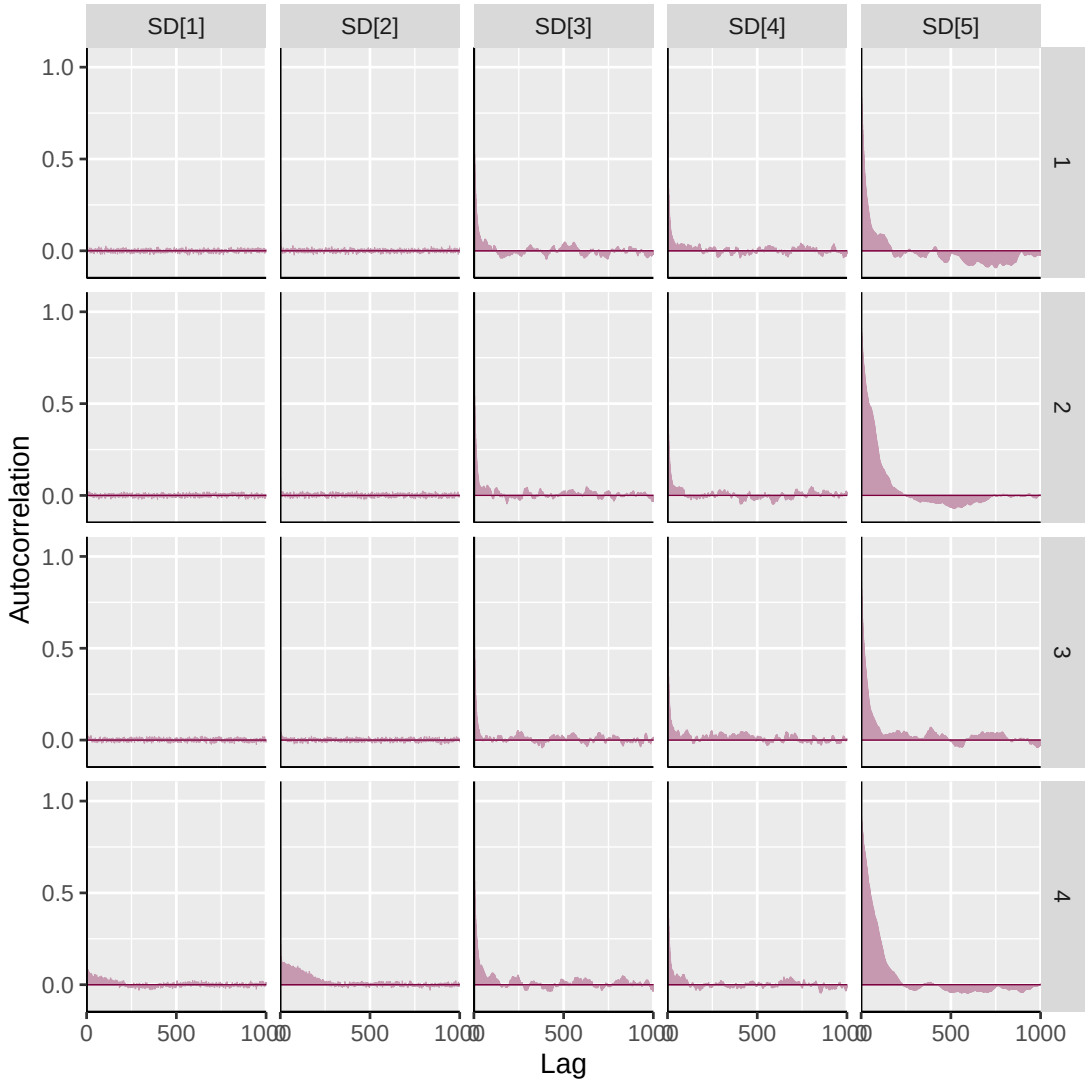


Figure A.4: Autocorrelation plots of standard deviations δ_t by chain



A.1.3 Marginal log odds ratios mLOR_{ij}

Figure A.5: Trace plots of log odds ratios mLOR_{ij} by chain

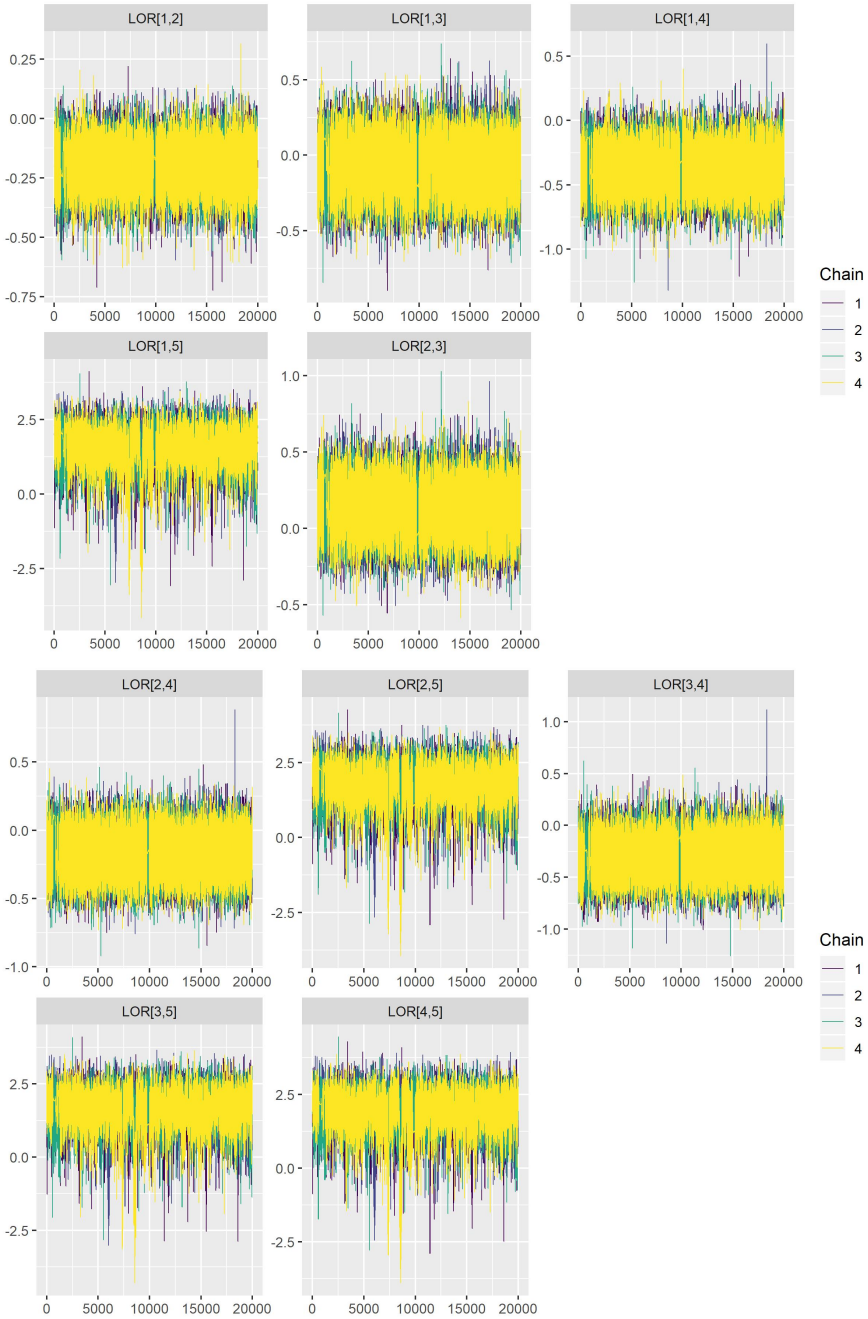
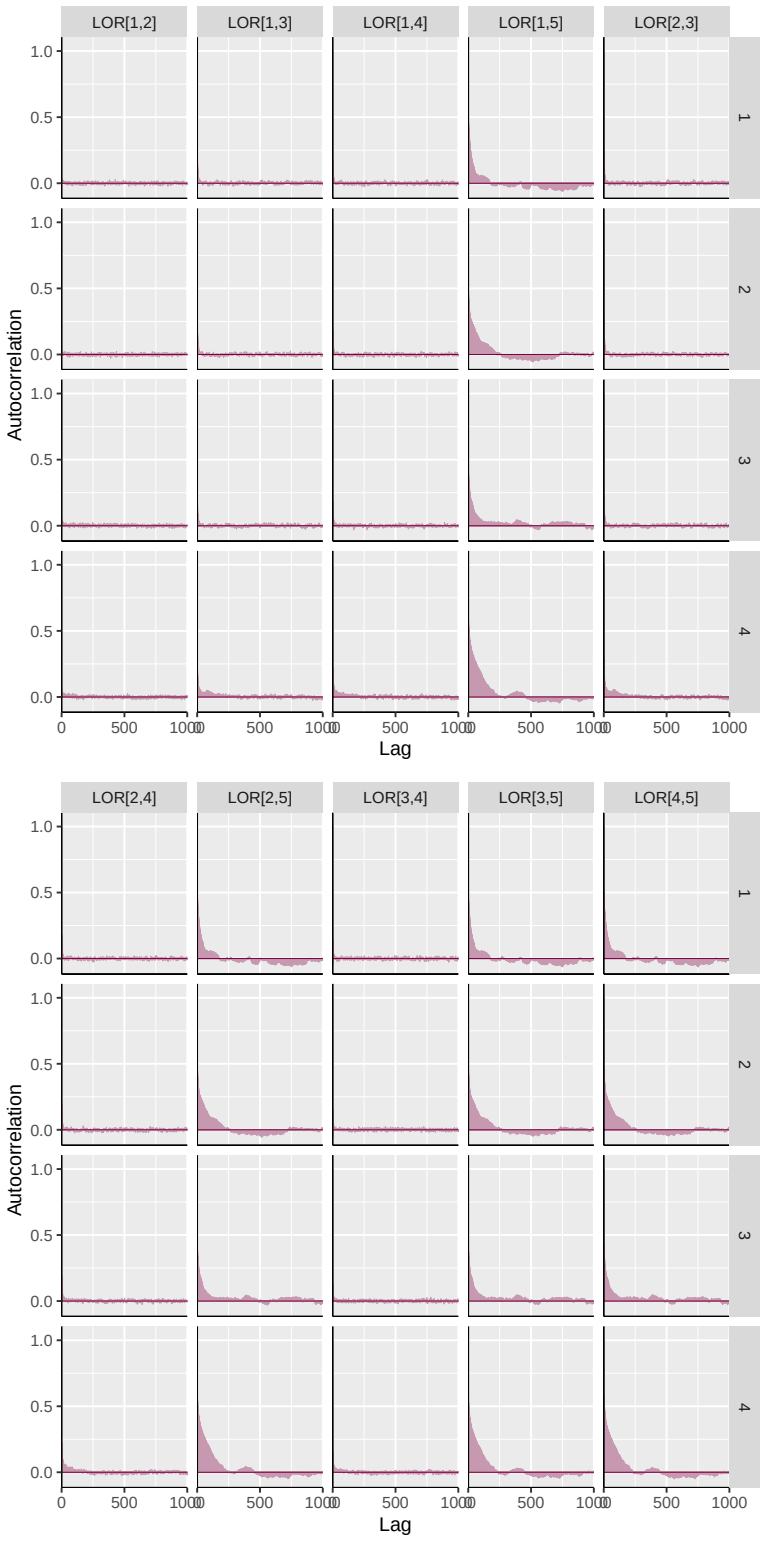


Figure A.6: Autocorrelation plots of log odds ratios mLOR_{ij} by chain



A.2 Results for hierarchical inverse-gamma (HIG) method

We further considered another shrinkage prior mentioned in Section 3.6: hierarchical inverse-gamma prior. We compared the following five methods in the case study and simulation studies.

- Model 1: The inverse-Wishart (IW) prior, the conjugate prior for the multivariate normal. The prior for the covariance matrix $\mathbf{\Sigma}$ 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 $\mathbf{I}$.

- 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 $U(-\frac{1}{T-1}, 1)$ 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 the HHC prior 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 5: The hierarchical inverse-gamma (HIG) method. It shares the same setting with Model 2 except the HIG prior HIG$(0, 1)$ is used for δ_t.

Table A.1: Organized inpatient care for stroke data: Comparing posterior median and 95% credible intervals under 5 models (IW, UV, HHC, HIG and EV); mLOR_{ij} compares the i^{th} and j^{th} treatment, absolute risk of events for the t^{th} treatment (p_t), and standard deviation of the t^{th} treatment (δ_t). Treatment labels: 1) stroke ward, 2) general medical ward, 3) mixed rehabilitation ward, 4) mobile stroke team, and 5) acute (semi-intensive) ward.

	Point Estimate (95% Credible Interval)				
Parameter	IW	UV	HHC	HIG	EV
mLOR_{12}	−0.23 (−0.50, 0.03)	−0.19 (−0.39, −0.03)	−0.20 (−0.38, −0.03)	−0.21 (−0.38, −0.04)	−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)	−0.11 (−0.42, 0.20)	−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)	−0.39 (−0.70, −0.08)	−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)	1.27 (0.01, 2.30)	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)	0.10 (−0.20, 0.39)	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)	−0.18 (−0.46, 0.11)	−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)	1.48 (0.21, 2.50)	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)	−0.28 (−0.63, 0.09)	−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)	1.38 (0.11, 2.41)	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)	1.66 (0.37, 2.69)	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)	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)	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)	0.24 (0.19, 0.30)	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)	0.29 (0.24, 0.36)	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)	0.07 (0.03, 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)	0.70 (0.51, 0.99)	.
δ_2	0.69 (0.50, 0.97)	0.74 (0.52, 1.04)	0.68 (0.49, 0.96)	0.67 (0.49, 0.95)	.
δ_3	0.51 (0.31, 0.91)	0.35 (0.08, 0.81)	0.28 (0.04, 0.67)	0.39 (0.18, 0.74)	.
δ_4	0.52 (0.32, 0.99)	0.39 (0.12, 0.86)	0.31 (0.08, 0.68)	0.39 (0.19, 0.74)	.
δ_5	0.60 (0.33, 1.37)	0.54 (0.07, 4.20)	0.23 (0.01, 1.43)	0.44 (0.17, 1.34)	.
DIC	99.99	93.66	90.27	89.94	94.53
$\bar{D}$	53.93	56.00	54.50	54.41	58.68
p_D	46.05	37.66	35.77	35.53	35.86
WAIC	6742.95	6741.44	6738.99	6739.40	6747.61

Table A.2: Simulation results for data generated under scenario I (heterogeneous variance) with the four different missingness settings (MCAR1, MCAR2, MAR1, MAR2), specifically 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 5 different priors (IW, HHC, HIG, UV, and EV). Specifically as an example, the table entry in column $\text{Bias}_{\tilde{\mu}}$ and row cLOR_{ij} was defined as $\sum_{i\neq j} |\text{Bias}_{\tilde{\mu}}(\text{cLOR}_{ij})|$. Similarly, the table entry in column CP and row cLOR_{ij} was defined as $\sum_{i\neq j}(0.95 - \text{CP}(\text{cLOR}_{ij}))_+$.

		IW			HHC			HIG			UV			EV		
Parameter	Truth	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP
MCAR1																
cLOR_{ij}	.	0.50	2.58	0.00	0.32	2.48	0.01	0.46	2.44	0.01	0.44	2.79	0.00	0.75	2.69	0.00
mLOR_{ij}	.	0.80	2.22	0.00	0.07	2.31	0.02	0.46	2.16	0.00	1.11	2.59	0.00	1.69	2.34	0.00
p_t	.	0.02	0.00	0.02	0.02	0.00	0.03	0.02	0.00	0.03	0.07	0.01	0.00	0.06	0.01	0.05
δ_t	.	0.90	0.34	1.22	0.15	0.49	0.02	0.62	0.35	0.74	1.29	1.21	0.12	1.97	1.09	3.31
mLOR_{34}	−1.40	−0.04	0.15	0.99	0.01	0.16	0.97	−0.02	0.14	0.98	0.15	0.20	0.99	−0.04	0.14	0.99
p_5	0.06	0.00	0.00	0.96	0.00	0.00	0.95	0.00	0.00	0.95	0.01	0.00	0.96	0.01	0.00	0.97
ρ_{35}	0.50	−0.45	0.22	1.00	−0.01	0.04	0.98	−0.04	0.04	0.98	0.02	0.04	0.98	−0.11	0.05	0.97
MCAR2																
cLOR_{ij}	.	0.68	5.34	0.02	0.64	5.44	0.00	0.57	4.75	0.01	0.78	6.63	0.00	0.87	4.76	0.01
mLOR_{ij}	.	1.02	4.46	0.00	0.53	4.43	0.00	0.91	4.12	0.00	2.82	4.58	0.00	1.90	4.10	0.10
p_t	.	0.04	0.01	0.02	0.04	0.01	0.01	0.03	0.01	0.01	0.15	0.02	0.00	0.07	0.01	0.07
δ_t	.	0.93	0.31	1.44	0.19	1.06	0.00	0.64	0.35	0.36	3.22	3.71	0.11	1.82	0.95	3.23
mLOR_{34}	−1.40	−0.10	0.35	0.99	−0.05	0.34	0.99	−0.10	0.33	0.99	0.20	0.35	1.00	−0.10	0.32	0.99
p_5	0.06	0.00	0.00	0.98	0.01	0.00	0.96	0.00	0.00	0.96	0.03	0.00	0.98	0.01	0.00	0.98
ρ_{35}	0.50	−0.49	0.24	1.00	−0.03	0.04	1.00	−0.08	0.05	0.98	−0.03	0.04	1.00	−0.21	0.07	0.94
MAR1																
cLOR_{ij}	.	2.78	3.61	0.00	0.91	3.42	0.00	2.26	3.45	0.00	7.02	10.86	0.03	6.99	8.10	0.73
mLOR_{ij}	.	2.70	2.90	0.00	0.60	2.44	0.00	1.50	2.46	0.01	3.75	4.48	0.05	5.55	5.40	0.70
p_t	.	0.06	0.01	0.01	0.03	0.01	0.01	0.05	0.01	0.01	0.17	0.02	0.07	0.19	0.03	0.19
δ_t	.	1.02	0.41	1.25	0.30	0.61	0.01	0.78	0.45	0.83	2.72	3.09	0.33	2.18	1.33	3.35
mLOR_{34}	−1.40	0.42	0.39	0.99	0.02	0.22	0.99	−0.22	0.28	0.99	−0.44	0.48	0.99	−0.90	1.14	0.85
p_5	0.06	0.02	0.00	0.98	0.01	0.00	0.97	0.00	0.00	0.98	−0.00	0.00	0.99	−0.01	0.00	0.96
ρ_{35}	0.50	−0.49	0.25	1.00	0.01	0.04	1.00	0.06	0.05	1.00	0.29	0.11	0.94	0.18	0.08	0.91
MAR2																
cLOR_{ij}	.	4.81	7.17	0.00	0.80	6.41	0.00	1.63	5.68	0.02	10.52	21.84	0.00	7.50	11.34	0.38
mLOR_{ij}	.	4.57	5.92	0.00	0.99	4.93	0.01	1.79	4.76	0.00	6.75	9.04	0.00	6.46	8.40	0.48
p_t	.	0.09	0.01	0.01	0.04	0.01	0.01	0.06	0.01	0.01	0.33	0.06	0.01	0.24	0.04	0.11
δ_t	.	1.00	0.33	1.58	0.19	1.07	0.01	0.68	0.37	0.43	3.90	4.90	0.15	2.02	1.16	3.24
mLOR_{34}	−1.40	0.60	0.65	1.00	0.01	0.35	1.00	−0.24	0.45	0.99	−0.84	1.15	1.00	−1.02	1.61	0.90
p_5	0.06	0.03	0.00	0.99	0.01	0.00	0.99	0.01	0.00	0.98	0.00	0.00	1.00	−0.01	0.00	0.95
ρ_{35}	0.50	−0.50	0.25	1.00	−0.00	0.03	1.00	0.04	0.05	1.00	0.18	0.06	1.00	0.14	0.06	0.95

Table A.3: Simulation results for data generated under scenario II (homogeneous variance) with the four different missingness settings (MCAR1, MCAR2, MAR1, MAR2), specifically 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 5 different priors (IW, HHC, HIG, UV, and EV). Specifically as an example, the table entry in column $\text{Bias}_{\tilde{\mu}}$ and row cLOR_{ij} was defined as $\sum_{i\neq j} |\text{Bias}_{\tilde{\mu}}(\text{cLOR}_{ij})|$. Similarly, the table entry in column CP and row cLOR_{ij} was defined as $\sum_{i\neq j}(0.95 - \text{CP}(\text{cLOR}_{ij}))_+$.

		IW			HHC			HIG			UV			EV		
Parameter	Truth	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP	$\text{Bias}_{\tilde{\mu}}$	$\text{MSE}_{\tilde{\mu}}$	CP
MCAR1																
cLOR_{ij}	.	0.37	2.04	0.00	0.22	1.95	0.00	0.28	1.89	0.00	0.39	2.22	0.00	0.23	1.80	0.01
mLOR_{ij}	.	0.11	1.74	0.00	0.08	1.80	0.00	0.09	1.67	0.00	0.88	2.07	0.00	0.18	1.64	0.01
p_t	.	0.02	0.00	0.00	0.02	0.00	0.02	0.02	0.00	0.01	0.06	0.00	0.00	0.01	0.00	0.02
δ_t	.	0.51	0.09	0.00	0.10	0.36	0.00	0.16	0.15	0.00	1.12	0.97	0.11	0.10	0.05	0.00
mLOR_{34}	−1.44	−0.00	0.14	0.99	−0.00	0.14	0.97	0.00	0.13	0.98	0.09	0.16	0.99	−0.02	0.13	0.95
p_5	0.05	0.00	0.00	0.98	0.00	0.00	0.96	0.00	0.00	0.97	0.01	0.00	0.97	0.00	0.00	0.96
ρ_{35}	0.50	−0.47	0.23	1.00	0.00	0.04	0.99	−0.01	0.05	0.98	0.03	0.05	0.99	−0.04	0.05	0.97
MCAR2																
cLOR_{ij}	.	0.60	3.89	0.00	0.53	3.76	0.00	0.55	3.55	0.00	0.70	5.15	0.00	0.50	3.39	0.19
mLOR_{ij}	.	0.28	3.30	0.00	0.35	3.29	0.00	0.36	3.02	0.00	3.02	4.09	0.00	0.44	3.08	0.17
p_t	.	0.03	0.01	0.00	0.03	0.01	0.02	0.03	0.01	0.01	0.15	0.02	0.00	0.02	0.01	0.05
δ_t	.	0.62	0.11	0.00	0.07	0.58	0.03	0.30	0.17	0.00	3.25	3.65	0.11	0.08	0.05	0.04
mLOR_{34}	−1.44	−0.03	0.35	0.99	−0.03	0.33	1.00	−0.03	0.31	0.99	0.22	0.36	1.00	−0.07	0.32	0.93
p_5	0.05	0.00	0.00	0.98	0.00	0.00	0.96	0.00	0.00	0.98	0.03	0.00	0.98	0.00	0.00	0.93
ρ_{35}	0.5	−0.50	0.25	1.00	−0.00	0.05	0.99	−0.01	0.05	0.99	0.01	0.05	0.99	−0.03	0.05	0.98
MAR1																
cLOR_{ij}	.	3.24	2.89	0.00	0.17	2.40	0.00	0.58	2.22	0.00	5.56	7.32	0.00	0.40	1.97	0.02
mLOR_{ij}	.	3.30	2.65	0.00	0.43	2.07	0.00	0.36	1.89	0.00	3.19	3.70	0.01	0.42	1.78	0.01
p_t	.	0.07	0.00	0.00	0.02	0.01	0.02	0.02	0.00	0.01	0.15	0.02	0.03	0.01	0.00	0.02
δ_t	.	0.48	0.08	0.00	0.07	0.40	0.00	0.25	0.16	0.00	2.07	2.02	0.29	0.08	0.04	0.00
mLOR_{34}	−1.44	0.53	0.41	0.97	0.05	0.20	0.99	−0.04	0.19	1.00	−0.48	0.53	0.99	0.07	0.16	0.98
p_5	0.05	0.02	0.00	0.97	0.01	0.00	0.97	0.00	0.00	0.99	−0.00	0.00	1.00	0.00	0.00	0.97
ρ_{35}	0.50	−0.50	0.25	1.00	−0.02	0.04	1.00	0.01	0.04	1.00	0.28	0.10	0.98	−0.08	0.05	1.00
MAR2																
cLOR_{ij}	.	4.62	5.04	0.00	0.37	4.38	0.00	0.59	3.64	0.00	8.38	15.75	0.00	0.58	3.22	0.00
mLOR_{ij}	.	4.44	4.45	0.00	0.71	3.61	0.00	0.31	3.04	0.00	5.30	7.03	0.00	0.59	2.92	0.01
p_t	.	0.09	0.01	0.00	0.02	0.01	0.02	0.03	0.01	0.00	0.28	0.05	0.00	0.01	0.01	0.01
δ_t	.	0.60	0.10	0.00	0.07	0.57	0.02	0.36	0.17	0.00	3.77	4.49	0.13	0.08	0.05	0.00
mLOR_{34}	−1.44	0.71	0.74	0.99	0.10	0.39	1.00	−0.04	0.34	1.00	−0.62	0.91	1.00	0.07	0.30	0.96
p_5	0.05	0.03	0.00	0.99	0.01	0.00	0.99	0.00	0.00	0.99	0.01	0.00	1.00	0.00	0.00	0.96
ρ_{35}	0.50	−0.50	0.25	1.00	−0.02	0.04	1.00	−0.01	0.05	1.00	0.12	0.06	0.99	−0.06	0.05	0.99

Appendix B

Proofs of Theorems

B.1 Definition of restricted Wishart distribution and its properties

In this appendix, we want to show the equivalence of restricted Wishart distribution and LKJ distribution [43] in some sense. We start with the definition of Wishart distribution. As we know, if $T \times 1$ vectors $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_K$ are independent random variables that are multivariate normally distributed with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}$, and we have a $T \times K$ matrix $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_K)$, then the T-dimensional positive definite matrix $\mathbf{S} = \boldsymbol{\nu}\boldsymbol{\nu}'$ ($\boldsymbol{\nu}'$ is the transpose of $\boldsymbol{\nu}$) follows the Wishart distribution with degree of freedom $K > T-1$ and positive definite scale matrix $\boldsymbol{\Sigma}$:

$$\mathbf{S} \sim \mathrm{W}_T(K, \boldsymbol{\Sigma}) \equiv \{2^{\frac{1}{2}KT}\Gamma_T(\frac{1}{2}K)\}^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}K}|\mathbf{S}|^{\frac{1}{2}(K-T-1)}\exp(-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{S})) \qquad \text{(B.1)}$$

where $|\bullet|$ is the determinant, tr is the trace, and Γ_T is the multivariate (T-variate here) gamma function. After that, we can formally define the restricted Wishart distribution based on the Wishart distribution:

Definition 1. A positive definite correlation matrix $\mathbf{R}$ follows the restricted Wishart distribution $\mathrm{RW}_T(m)$, if $\mathbf{R} = \boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta}$, where $\boldsymbol{\Sigma} \sim \mathrm{W}_T(m, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi}$ being diagonal matrix with diagonal entries $\psi_{11}, \psi_{22}, \ldots, \psi_{TT}$, $\boldsymbol{\Delta}$ is a diagonal matrix with the i^{th} diagonal element $\sigma_{ii}^{-1/2}$ and σ_{ii} is the i^{th} diagonal element of $\boldsymbol{\Sigma}$.

Theorem 1. If R follows the restricted Wishart distribution $\mathrm{RW}_T(m)$, then

(a) the joint distribution of $\mathbf{R}$ is:

$$f(\mathbf{R}) = b_T(m)|\mathbf{R}|^{\frac{m-T-1}{2}} \tag{B.2}$$

where

$$b_T(m) = \frac{\Gamma^T(\frac{m}{2})}{\Gamma_T(\frac{m}{2})} \tag{B.3}$$

is the normalizing constant. When $m = T+1$, R becomes a jointly uniform distribution on a compact subspace of the $T(T-1)/2$ dimensional hypercube $[-1,1]^{T(T-1)/2}$.
(b) this normalizing constant $b_T(m)$ matches the following alternative formula given by Lewandowski et al. [43]:

$$c_T = 2^{\sum_{k=1}^{T-1}(m-k-1)(k-T)} \times \prod_{k=1}^{T-1}[B(\frac{m-k}{2}, \frac{m-k}{2})]^{k-T} \tag{B.4}$$

where $B(\bullet, \bullet)$ is the beta function.
(c) the marginal distribution of each element ρ_{ij} $(i \neq j)$ in matrix $\mathbf{R}$ is

$$f(\rho_{ij}) = \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m-1}{2})\Gamma(\frac{1}{2})}(1-\rho_{ij}^2)^{\frac{m-3}{2}}, \quad -1 \leq \rho_{ij} \leq 1 \tag{B.5}$$

which is exactly the $Beta(\frac{m-1}{2}, \frac{m-1}{2})$ distribution on $[-1,1]$, and will be $Beta(\frac{T}{2}, \frac{T}{2})$ when $f(\mathbf{R}) = b_T(T+1)$.

Proof of Theorem 1. (a). We first calculate the Jacobian matrix of the transformation $\mathbf{\Sigma} \to (\sigma_{11}, \ldots, \sigma_{TT}, \mathbf{R})$

$$\mathbf{J}_1 = \mathbf{J}(\sigma_{11} = \sigma_{11}, \ldots, \sigma_{TT} = \sigma_{TT}, \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}) = \left[\begin{array}{c|c} \mathbf{I}_T & 0 \\ \hline * & \mathbf{C} \end{array}\right] \tag{B.6}$$

where σ_{ij} is the entry of matrix $\mathbf{\Sigma}$, $\mathbf{I}_T$ is the T-dimensional identity matrix, $\mathbf{C}$ is the $T(T-1)/2$ dimensional diagonal matrix with entries $1/\sqrt{\sigma_{ii}\sigma_{jj}}$ $(i \neq j)$. Since J_1 is a lower triangular matrix, its determinant $|\mathbf{J}_1|$ equals $\prod_{i=1}^{T}\sigma_{ii}^{-(T-1)/2}$ and sub-matrix in (B.6) given as $*$ need not be derived. Then, given $\mathbf{\Sigma} \sim \mathrm{W}_T(m, \mathbf{\Psi})$, we can derive the joint distribution of $\mathbf{R}$ and $(\sigma_{11}, \ldots, \sigma_{TT})$ is

$$\begin{aligned} f(\mathbf{R}, \sigma_{11}, \ldots, \sigma_{TT}) &= f(\mathbf{\Sigma})|\mathbf{J}_1|^{-1} \\ &= \frac{\Gamma^T(\frac{m}{2})|\mathbf{R}|^{\frac{m-T-1}{2}}}{\Gamma_T(\frac{m}{2})} \prod_{i=1}^{T}\{\frac{\sigma_{ii}^{\frac{m}{2}-1}\exp(-\frac{\sigma_{ii}}{\psi_{ii}})}{2^{\frac{m}{2}}\psi_{ii}^{\frac{m}{2}}\Gamma(\frac{m}{2})}\} \end{aligned} \tag{B.7}$$

Clearly, $\frac{\sigma_{ii}}{\psi_{ii}}, i = 1, \ldots, T$ are independently distributed as chi-square with m degrees of freedom. Also, the density of $\boldsymbol{P}$ is

$$f(\mathbf{R}) = \frac{\Gamma^T(\frac{m}{2})}{\Gamma_T(\frac{m}{2})}|\mathbf{R}|^{\frac{m-T-1}{2}} \tag{B.8}$$

(b). We just need to prove function $f(T, m) = b_T(m)/c_T$ equals 1 for any integer T (≥ 2) and $m > T-1$. Given $\Gamma_T(\frac{m}{2}) = \pi^{T(T-1)/4}\prod_{k=1}^{T}\Gamma(\frac{m}{2} + \frac{1-k}{2})$ (James [149], p. 483), we can simplify $f(T, m)$ as

$$\begin{aligned} f(T,m) &= \frac{\Gamma^T(\frac{m}{2})}{\Gamma_T(\frac{m}{2})}\prod_{k=1}^{T-1}[B(\frac{m-k}{2}, \frac{m-k}{2})]^{T-k} \times 2^{\sum_{k=1}^{T-1}(m-k-1)(T-k)} \\ &= \frac{\Gamma^T(\frac{m}{2})}{\pi^{T(T-1)/4}\Gamma(\frac{m}{2})\prod_{k=1}^{T-1}\Gamma(\frac{m-k}{2})}\prod_{k=1}^{T-1}[\frac{\Gamma^2(\frac{m-k}{2})}{\Gamma(m-k)}]^{T-k} \times 2^{\sum_{k=1}^{T-1}(m-k-1)(T-k)} \\ &= \prod_{k=1}^{T-1}[2^{(m-k-1)(T-k)} \times \frac{\Gamma(\frac{m}{2})}{\pi^{T/4}}\frac{\Gamma^{2T-2k-1}(\frac{m-k}{2})}{\Gamma^{T-k}(m-k)}] \end{aligned} \tag{B.9}$$

We prove that (B.9) equals 1 using mathematical induction. Start with $T = 2$, then for any $m > 1$, $f(2, m)$ reduces to

$$2^{m-2}\pi^{-\frac{1}{2}}\frac{\Gamma(\frac{m}{2})\Gamma(\frac{m-1}{2})}{\Gamma(m-1)} \equiv 1 \tag{B.10}$$

which is known as a special case ($k = 1$) of the identity $\Gamma(m-k) \equiv 2^{m-k-1}\pi^{-1/2}\Gamma(\frac{m-k}{2})\Gamma(\frac{m-k+1}{2})$ (Abramowitz and Stegun [150], p. 483) . Assume $f(T, m) = 1$ holds for $T = t$. It must then be shown that $f(t+1, m) = 1$, $\forall m > T-1$, where

$$\begin{aligned} f(t+1,m) &= \prod_{k=1}^{t}[2^{(m-k-1)(t+1-k)} \times \frac{\Gamma(\frac{m}{2})}{\pi^{(t+1)/4}}\frac{\Gamma^{2t-2k+1}(\frac{m-k}{2})}{\Gamma^{t+1-k}(m-k)}] \\ &= [2^{m-t-1}\frac{\Gamma(\frac{m}{2})}{\pi^{(t+1)/4}}\frac{\Gamma(\frac{m-t}{2})}{\Gamma(m-t)}] \times \prod_{k=1}^{t-1}[2^{(m-k-1)(t+1-k)} \times \frac{\Gamma(\frac{m}{2})}{\pi^{(t+1)/4}}\frac{\Gamma^{2t-2k+1}(\frac{m-k}{2})}{\Gamma^{t+1-k}(m-k)}] \\ &= [2^{m-t-1}\frac{\Gamma(\frac{m}{2})}{\pi^{(t+1)/4}}\frac{\Gamma(\frac{m-t}{2})}{\Gamma(m-t)}] \times f(t,m) \times \prod_{k=1}^{t-1}[2^{m-k-1}\frac{1}{\pi^{1/4}}\frac{\Gamma^2(\frac{m-k}{2})}{\Gamma(m-k)}] \end{aligned} \tag{B.11}$$

Using the same identity $\Gamma(m-k) \equiv 2^{m-k-1}\pi^{-1/2}\Gamma(\frac{m-k}{2})\Gamma(\frac{m-k+1}{2})$ multiple times, $f(t+1,m)$ is

$$\begin{aligned} f(t+1,m) &= [\frac{1}{\pi^{(t-1)/4}}\frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m-t+1}{2})}] \times 1 \times \prod_{k=1}^{t-1}[\frac{1}{\pi^{-1/4}}\frac{\Gamma(\frac{m-k}{2})}{\Gamma(\frac{m-k+1}{2})}] \\ &= \prod_{k=1}^{t}[\frac{\Gamma(\frac{m-k+1}{2})}{\Gamma(\frac{m-k+1}{2})}] \\ &= 1 \end{aligned} \tag{B.12}$$

(c). According to Eaton [151] (p. 256), if $\mathbf{\Sigma} \sim \mathrm{W}_T(m, \mathbf{\Psi})$, then any $T_1 \times T_1$ principal sub-matrix $\mathbf{\Sigma}_1$ of $\mathbf{\Sigma}$ has following distribution

$$\mathbf{\Sigma}_1 \sim W_{T_1}(m, \mathbf{\Psi}_1) \tag{B.13}$$

with $\mathbf{\Psi}_1$ being the $T_1 \times T_1$ principal sub-matrix of $\mathbf{\Psi}$. By applying it to the (B.2), any $T_1 \times T_1$ principal sub-matrix $\mathbf{R}_1$ of $\mathbf{R}$ has density function $b_{T_1}(m)|\mathbf{R}_1|^{\frac{m-T_1-1}{2}}$. Simply let T_1=2, we could obtain the marginal distribution of ρ_{ij}

$$f(\rho_{ij}) = \frac{\Gamma^2(\frac{m}{2})}{\Gamma_2(\frac{m}{2})}(1-\rho_{ij}^2)^{\frac{m-3}{2}}, \quad -1 \le \rho_{ij} \le 1 \tag{B.14}$$

□