HHS Public Access

Author manuscript

Ann Appl Stat. Author manuscript; available in PMC 2022 August 26.

Published in final edited form as:

Ann Appl Stat. 2021 December ; 15(4): 1767–1787. doi:10.1214/21-aoas1469.

BRIDGING RANDOMIZED CONTROLLED TRIALS AND SINGLE-ARM TRIALS USING COMMENSURATE PRIORS IN ARM-BASED NETWORK META-ANALYSIS

Zhenxun Wang[1], **Lifeng Lin**[2], **Thomas Murray**[1], **James S. Hodges**[1], **Haitao Chu**[1]

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

[2]Department of Statistics, Florida State University, Tallahassee, FL 32306, USA

Abstract

Network meta-analysis (NMA) is a powerful tool to compare multiple treatments directly and indirectly by combining and contrasting multiple independent clinical trials. Because many NMAs collect only a few eligible randomized controlled trials (RCTs), there is an urgent need to synthesize different sources of information, e.g., from both RCTs and single-arm trials. However, single-arm trials and RCTs may have different populations and quality, so that assuming they are exchangeable may be inappropriate. This article presents a novel method using a *commensurate prior on variance* (CPV) to borrow variance (rather than mean) information from single-arm trials 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 trials. The simulation results show that the CPV method provides efficient and robust estimation even when the two sources of information are moderately inconsistent.

Keywords and phrases:

Bayesian inference; commensurate prior; network meta-analysis; randomized controlled trial; single-arm trial

1. Introduction.

Meta-analyses and network meta-analyses (NMAs) are fundamental tools to quantitatively and rigorously assess efficacy and cost-effectiveness of interventions in evidence synthesis (Welton et al., 2012), analyzing many studies at the same time. While standard pairwise meta-analyses can compare only two treatments, NMA was developed to compare multiple (≥ 3) interventions simultaneously. Both contrast-based NMA (CB-NMA) (Lu and Ades, 2004, 2006, 2009) and arm-based NMA (AB-NMA) (Zhang et al., 2014; Hong et al., 2016a;

Supplementary Materials. The supplementary materials include the complete motivating dataset (Appendix A), additional details of the proposed methods (Appendix B), details of simulation results (Appendix C), additional simulation results (Appendix D), sensitivity analyses (Appendix E), case study by CB-NMA approaches (Appendix F). Data and sample R/nimble code can be found online in the Supporting Information section at the end of the article.

Zhang et al., 2017a) frameworks have been proposed, with the main difference lying in what they assume is exchangeable across studies: absolute treatment effects in AB-NMA, relative treatment effects (contrasts) in CB-NMA. Generally, randomized controlled trials (RCTs) with a blinded outcome assessment offer high-quality and reliable evidence for statistical analyses (Egger, Davey Smith and Altman, 2001) 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 RCTs (Kontopantelis, Springate and Reeves, 2013). It is challenging to select an appropriate method for meta-analysis with only a few RCTs (≤ 5), balancing statistical power and nominal coverage probability (Mathes and Kuss, 2018). 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 with an interquartile range of 1–4 (Nikolakopoulou et al., 2014; Wang et al., 2021). Because of this, variances of outcomes for individual treatments (absolute effects) in AB-NMA and variances for individual treatment comparisons (relative effects) in CB-NMA are difficult to estimate. Hence, it is natural to consider an extrapolation strategy in meta-analysis and NMA.

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 by incorporating historical controls when diseases are rare or patient populations are small (Chen et al., 2011; Hueber et al., 2012; Gamalo, Tiwari and LaVange, 2013; Gamalo-Siebers et al., 2017). While the history of borrowing external information in evidence synthesis dates back to the 1990s, when Begg and Pilote (1991) and Li and Begg (1994) 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. (2019) proposed methods for a meta-analyses to adaptively combine RCTs and single-arm trials, while Röver, Wandel and Friede (2018) used Bayesian model averaging to borrow adult evidence in pediatric meta-analysis, which generally has fewer trials available. In CB-NMA, an approach to incorporate single-arm trial data by using aggregate-level covariate matching has been proposed and discussed (Jaff et al., 2017; Schmitz et al., 2018; Leahy et al., 2019; Phillippo et al., 2020). In addition, Efthimiou et al. (2017) recently proposed approaches to combining randomized and non-randomized evidence in CB-NMA, while Thom et al. (2015) proposed a method to conduct indirect comparisons in an incomplete network by including single-arm observational studies. Turner et al. (2019) introduced four different informative priors for multiple heterogeneity variances in CB-NMA. There are also AB-NMA methods to synthesize aggregate and individual patient data (Hong, Fu and Carlin, 2018).

So far, however, very little attention has been paid to including single-arm trials in an AB-NMA, partly because there is an ongoing debate about CB-NMA versus AB-NMA (Dias and Ades, 2016; Hong et al., 2016b; White et al., 2019). The AB approach has the potential to estimate marginal absolute risks, which are necessary to calculate the incremental cost effectiveness ratio for comparing the cost-effectiveness of health care interventions. As mentioned above, however, scant information is so prevalent in NMA that it is difficult to estimate the standard deviations of treatment-specific effects across trials (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 (Wang et al., 2021). Hence, to provide better estimates it is necessary to develop methods that can incorporate extra evidence from single-arm trials in AB-NMA.

Several statistical methods have been developed for "information borrowing" using Bayesian methods, for instance, power priors (Chen and Ibrahim, 2000; Duan, Ye and Smith, 2005; Ibrahim et al., 2015), hierarchical commensurate priors (Hobbs et al., 2011; Hobbs, Sargent and Carlin, 2012; Murray, Hobbs and Carlin, 2015), and Bayesian model averaging (Schmidli et al., 2014; Kaizer, Koopmeiners and Hobbs, 2018; Kaizer, Hobbs and Koopmeiners, 2018). Motivated by these methods, we propose commensurate priors to adaptively incorporate *variance* information from single-arm trials into an AB-NMA. Although an AB-NMA naturally incorporates single-arm trials (Lin, Chu and Hodges, 2016), current methods do not explicitly account for the possibly lower quality of single-arm trials. Our new method, by contrast, has the advantage of downweighting single-arm trials when they appear to be inconsistent with two- or multi-arm RCTs in an NMA.

The rest of this article is organized as follows. Section 2 describes a motivating example, an NMA comparing safety of different immune checkpoint inhibitors for treating cancer. Section 3 introduces commensurate priors to combine RCTs and single-arm trials in an AB-NMA. Section 4 presents results from applying our method to the motivating example, followed by simulation studies in Section 5, comparing the performance of different commensurate priors. Section 6 summarizes our findings and discusses future research.

2. Motivating example.

Immune checkpoint inhibitors (ICIs) have recently emerged as a breakthrough in treating more than 14 cancers including melanoma, Hodgkin lymphoma, non-small cell lung cancer, and others (Johnson, Chandra and Sosman, 2018). To investigate the safety of ICIs, Xu et al. (2018) conducted a systematic review and NMA mainly on five ICIs, ipilimumab, tremelimumab, nivolumab, pembrolizumab, and atezolizumab. Only 3 out of 31 RCTs had an atezolizumab arm, making parameters related to this ICI (e.g., variance, absolute risk, and relative risk) difficult to estimate. Fortunately, Xu et al. (2018) identified not only the 31 RCTs but also found 36 single-arm trials. We collected all available data from these RCTs and single-arm trials for our analysis.

Table A1 in Appendix A presents a cleaned dataset of 27 RCTs and 28 single-arm trials 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. Figure 1 intuitively shows the need to borrow information from single-arm trials to potentially improve estimation. For example, only 4 and 3 RCTs contain ipilimumab (high dose) and atezolizumab respectively, causing difficulty in estimating the variances of these two treatment-specific log-odds. Nevertheless,

with the additional information from single-arm trials (3 for each of these two ICIs), we may be able to overcome this problem.

3. Statistical methods.

3.1. Notation.

Assume an NMA has K RCTs comparing a total of T treatments. Let $\mathscr{A}_k$ ($k = 1, \ldots, K$) be the subset of treatments in the k^{th} trial. For most RCTs, the number of treatments in $\mathscr{A}_k$, denoted by $|\mathscr{A}_k|$, is 2 or 3. Let $\mathscr{D}_k$ be the data observed in the k^{th} RCT. For NMAs with a dichotomous outcome, $\mathscr{D}_k = \{(r_{kt}, n_{kt}), t \in \mathscr{A}_k\}$, 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 single-arm trials. Let $\mathscr{D}_j^{\text{S}} (j = 1, \ldots, J)$ be the data collected in the j^{th} single-arm trial; $\mathscr{D}_j^{\text{S}} = \left\{\left(r_{jt}^{\text{S}}, n_{jt}^{\text{S}}\right), t \in \mathscr{A}_j^{\text{S}}\right\}$, where $\mathscr{A}_j^{\text{S}}$ includes only one treatment with $\left|\mathscr{A}_j^{\text{S}}\right| = 1$, and r_{jt}^{S} and n_{jt}^{S} are the numbers of events and participants for the t^{th} treatment in the j^{th} single-arm trial. We further define B_t to be the number of RCTs containing the t^{th} treatment, and B_t^{S} to be the number of single-arm trials containing the t^{th} treatment. For instance, as shown in Figure 1, 9 RCTs and 11 single-arm trials contain nivolumab, so $B_1 = 9$ and $B_1^{\text{S}} = 11$.

3.2. Arm-based network meta-analysis and model for single-arm trials.

This subsection briefly introduces the AB-NMA (Zhang et al., 2014; Hong et al., 2016a) and the model for single-arm trials, focusing on binary outcomes. The AB-NMA model for RCTs is:

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

where p_{kt} is the probability of an event (i.e., absolute risk) for the t^{th} treatment in the k^{th} trial 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 $|\mathscr{A}_k|$ treatments t are actually observed in trial k. The vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)'$ contains the overall logit event probability for each treatment. If we denote the between-trial standard deviation of treatment t by σ_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 σ_t being its t^{th} diagonal element. We further define $\mu^{\text{RCT}} = (\mu_1, \ldots, \mu_T)'$, and $\boldsymbol{\sigma}^{\text{RCT}} = (\sigma_1, \ldots, \sigma_T)'$. This model is exactly the same as Model 4 in White et al. (2019) though with slightly different notation. We call this original method *no borrowing* (NB) because it does not incorporate any information from single-arm trials.

Similarly, the model for single-arm trials is:

$$
\begin{aligned}
&r^{\mathrm{s}}_{jt} \sim \text{Binomial}\left(n^{\mathrm{s}}_{jt}, p^{\mathrm{s}}_{jt}\right), t \in \mathscr{A}^{\mathrm{s}}_{j}, j = 1, \ldots, J; \\
&\text{logit}\left(p^{\mathrm{s}}_{jt}\right) = \theta^{\mathrm{s}}_{jt}; \\
&\theta^{\mathrm{s}}_{jt} \sim N\left(\mu^{\mathrm{s}}_{t}, \left(\sigma^{\mathrm{s}}_{t}\right)^{2}\right),
\end{aligned}
\tag{2}
$$

where p^{s}_{jt} is the probability of an event for the j^{th} single-arm trial, μ^{s}_{t} represents the overall fixed effect of treatment t from single-arm trials, and σ^{s}_{t} is the standard deviation of the t^{th} treatment for single-arm trials. Unlike Equation (1), which contains some latent variables corresponding to unobserved treatment arms, the variables in Equation (2) all correspond to observed treatment arms. We further define $\boldsymbol{\mu}^{\mathrm{s}} = \left(\mu^{\mathrm{s}}_{1}, \ldots, \mu^{\mathrm{s}}_{T}\right)'$ and $\boldsymbol{\sigma}^{\mathrm{s}} = \left(\sigma^{\mathrm{s}}_{1}, \ldots, \sigma^{\mathrm{s}}_{T}\right)'$.

3.3. Connecting NMA and single-arm trials.

Based on the models above for an NMA and single-arm trials, we consider several methods to adaptively integrate information from the single-arm trials into the NMA.

3.3.1. Existing methods: full borrowing.—The AB-NMA model in Equation (1) could naturally incorporate information from single-arm trials about means and variances by assuming $\mu_t = \mu^{\mathrm{s}}_{t}$ and $\sigma_t = \sigma^{\mathrm{s}}_{t} (t = 1, \ldots, T)$. 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 trials by assuming $\sigma_t = \sigma^{\mathrm{s}}_{t}$ while $\mu_t \neq \mu^{\mathrm{s}}_{t}$. We call this method *fully borrowing on variances* (FBV). We could also borrow mean information only by assuming $\mu_t = \mu^{\mathrm{s}}_{t}$ while $\sigma_t \neq \sigma^{\mathrm{s}}_{t}$, but this article will not discuss this *fully borrowing on mean* (FBM) method in detail as it might be less useful in practice.

3.3.2. Commensurate prior on mean.—Although a full-borrowing approach naturally integrates single-arm trials, it may also cause large biases if the reliability of single-arm trials may be doubtful. Instead, a commensurate prior on the means, introduced by Hobbs, Sargent and Carlin (2012), is a simple, flexible way to borrow from and downweight single-arm trials:

$$
\mu_t \sim N\left(\mu^{\mathrm{s}}_{t}, \eta^{-1}\right); \tag{3}
$$

that is, μ_t has a normal prior with mean μ^{s}_{t} and precision η. The precision η characterizes how commensurate the two sources of information (μ_t and μ^{s}_{t}) are with each other. Hobbs, Sargent and Carlin (2012) proposed a "spike-and-slab" prior for η, but estimation of η is still difficult with this prior. Instead, Murray, Hobbs and Carlin (2015) proposed a modified commensurate prior:

$$
\begin{aligned}
&\mu_t \sim \left[N\left(\mu^{\mathrm{s}}_{t}, \left(\tau^{\mathrm{m}}_{t}\right)^{-1}\right)\right]^{1-\kappa^{\mathrm{m}}_{t}} \left[N\left(\mu^{\mathrm{s}}_{t}, \left(R^{\mathrm{m}}\right)^{-1}\right)\right]^{\kappa^{\mathrm{m}}_{t}}; \\
&\kappa^{\mathrm{m}}_{t} \sim \text{Bern}\left(p^{\mathrm{m}}\right) \text{ and } \tau^{\mathrm{m}}_{t} \sim U\left(s^{\mathrm{m}}_{1}, s^{\mathrm{m}}_{\mathrm{u}}\right), t = 1, \ldots, T,
\end{aligned}
\tag{4}
$$

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

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

$$\log(\sigma_t/\sigma_t^{\mathrm{s}}) = c_t;\ c_t \sim N\left(0, \eta^{-1}\right), \tag{5}$$

so the log of the standard deviation ratio follows a normal distribution with mean zero and precision η. Like Murray, Hobbs and Carlin (2015), we can modify this prior as follows:

$$\begin{aligned} &\log(\sigma_t) \sim \left[N\left(\log(\sigma_t^{\mathrm{s}}), \left(\tau_t^{\mathrm{v}}\right)^{-1}\right)\right]^{1-\kappa_t^{\mathrm{v}}} \left[N\left(\log(\sigma_t^{\mathrm{s}}), \left(R^{\mathrm{v}}\right)^{-1}\right)\right]^{\kappa_t^{\mathrm{v}}}; \\ &\kappa_t^{\mathrm{v}} \sim \mathrm{Bern}(p^{\mathrm{v}}) \text{ and } \tau_t^{\mathrm{v}} \sim U(s_{\mathrm{l}}^{\mathrm{v}}, s_{\mathrm{u}}^{\mathrm{v}}), t = 1, \ldots, T, \end{aligned} \tag{6}$$

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

3.3.4. Double commensurate prior.—We can borrow both mean and variance information adaptively by applying both the CPV and CPM methods in Equations (4) and (6). We call this method *commensurate prior on mean and variance* or *double commensurate prior* (DCP); it is an adaptively borrowing version of the FBMV method.

3.3.5. Summary of prior specifications and models.—Table 1 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, McCulloch and Meng (2000). 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 σ_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 $\mathbf{P}$ (Lin et al., 2017; Wang et al., 2020a): we assume all correlation coefficients ρ_{ij} are equal, i.e., $\rho_{ij} = \rho$ for any $i \ne j$, and assign a uniform prior $U\left(-\frac{1}{T-1}, 1\right)$ to

ρ so **P** is positive-definite. For models in which mean or variance information is not shared between the RCTs and single-arm trials 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 σ_t and σ_t^s. On the other hand, if information is shared fully or adaptively between the RCTs and single-arm trials, we assume $\mu_t = \mu_t^s$ and $\sigma_t = \sigma_t^s$ for fully borrowing, or for adaptively borrowing we follow Equations (4) and (6) with pre-specified values (0.5, 2500, 0, 2) for (p^m, R^m, s_1^m, s_u^m) and (p^v, R^v, s_1^v, s_u^v).

3.4. Likelihood and posterior estimation.

The likelihood functions are provided in Appendix B. We used NIMBLE (de Valpine et al., 2017) to fit the proposed models both for the real dataset on ICI safety 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 (Zeger, Liang and Albert, 1988):

$$\begin{aligned} &\mathrm{LOR}_{ij} = \mu_i - \mu_j; \\ &p_t = E[p_{kt} \mid \mu_t, \sigma_t] \approx \left[1 + \exp\left(-\mu_t / \sqrt{1 + \frac{256}{75\pi^2}\sigma_t^2}\right)\right]^{-1}. \end{aligned} \tag{7}$$

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

We chose NIMBLE for computations because it is much faster than JAGS (de Valpine, 2016). NIMBLE code, which is very similar to WinBUGS or JAGS code, is given online in the Supporting Information section.

3.5. Model comparison.

The deviance information criterion (DIC) (Spiegelhalter et al., 2002) and the logarithm of the pseudo marginal likelihood (LPML) (Geisser and Eddy, 1979; Hanson, Branscum and Johnson, 2011) are two popular criteria for comparing Bayesian models. Because estimates of single-arm trials are much less meaningful, we focus on the NMA part of the joint model. Appendix B describes procedures to estimate LPML. DIC can be calculated following the steps in Dias et al. (2013). 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 (Lunn et al., 2010).

4. Data analysis.

We applied the six models in Table 1 to the motivating example of the ICI data and compared the results. (We also applied two CB-NMA models to this dataset. Appendix F gives the results, which are discussed below in Section 6.) This dataset does not have single-arm trials for treatments 7 (2ICIs) and 8 (ICC), so the model settings are slightly different from those described above. To demonstrate the differences differences between

the model settings described above and used for this dataset, Figure 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 $t = 7, 8$, $\mu_t^S \sim N\left(0, 100^2\right)$ for $t = 1, \ldots, 6$, $\sigma_t \sim U(0, 5)$ for $t = 7, 8$, and $\sigma_t^S \sim U(0, 5)$ for $t = 1, \ldots, 6$.

To summarize rankings of the treatments in terms of safety, we use the surface under the cumulative ranking (SUCRA) proposed by Salanti, Ades and Ioannidis (2011). 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

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

where the 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.

Table 2 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 (σ_t), selected log odds ratios LOR_{ij}, LPML, and DIC. These models did not differ notably in LPML or DIC. Some differences, however, appear in the estimates and intervals.

Figure 3a is a forest plot of the posteriors for the standard deviations σ_t. Clearly, because of lack of information about treatments 3 (IPI high), 5 (ATE), and 7 (2ICIs), the estimates of σ_3, σ_5, and σ_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 trials for treatments IPI high and ATE, we may have better estimates for σ_3 and σ_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 trials differed, the FBV method had a much stronger effect on the posterior of σ_6 than the adaptive (CPV) method; the posterior median and 95% CrI of σ_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 σ_3 and σ_5 a bit by sharing mean information from single-arm trials.

The differences between methods in mean and variance estimates can affect the estimates of absolute risks, as shown in Figure 3b. The NB method yielded a much wider CrI for treatments IPI high dose and ATE than the CPV and FBV methods because 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 trials,

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 trials. 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 trials. 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, known as the plate plot (Wang et al., 2020b), visualizes the estimated log odds ratios (with results for NB shown above the diagonal and CPV below the diagonal) and SUCRAs (shown as a percent). 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 trials are potentially unreliable, the CPV method can provide better estimates than other methods that borrow mean information. Meanwhile, unlike the NB or FBV methods, the CPV method allows treatments with limited data to adaptively incorporate variance information from single-arm trials to improve estimates related to these treatments.

5. Simulation studies.

5.1. Simulation settings.

These simulation studies compared five methods (NB, CPV, FBV, CPM, and DCP) defined in Section 3. Each simulated dataset contained $K = 14$ RCTs, $J = 30$ single-arm trials, 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 trials were allocated to treatments with the pre-specified partition scheme of $B_1^S = 14$, $B_2^S = 5$, $B_3^S = 4$, $B_4^S = 4$, and $B_5^S = 3$. The number of participants in each single-arm trial 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 trials. 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 (1), with $\boldsymbol{\mu}^{\text{RCT}} = (\mu_1, \ldots, \mu_5)' = (-2, -3, -2.5, -2, -1.5)'$ 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) $\boldsymbol{\sigma}^{\text{RCT}} = (\sigma_1, \ldots, \sigma_5)' = (0.4, 1.0, 1.0, 0.3, 0.3)'$. In this complete dataset for the RCTs, each trial had 5 arms with $|\mathscr{A}_k| = 5$. To generate the 30 single-arm trials, we used Equation (2) with $\mu^S = \left(\mu_1^S, \ldots, \mu_5^S\right)' = (-2, -3, -2.5, -2, -1.5)'$ and $\sigma^S = \left(\sigma_1^S, \ldots, \sigma_5^S\right)' = (0.4, 1.0, 1.0, 0.3, 0.3)'$. In these single-arm trials, each trial only had one arm with $\left|\mathscr{A}_j^S\right| = 1$. The only difference between EM-EV and the other three scenarios was the pre-specified values for $\boldsymbol{\mu}^{\text{RCT}}$, $\boldsymbol{\sigma}^{\text{RCT}}$, $\boldsymbol{\mu}^S$, and $\boldsymbol{\sigma}^S$ as follows. The UM-EV (unequal mean and equal variance) scenario had $\boldsymbol{\mu}^{\text{RCT}} = (-2, -3, -2.5, -2, -1.5)'$, $\boldsymbol{\mu}^S = (-1, -2.5, -2, -2.5, -2)'$, and $\boldsymbol{\sigma}^{\text{RCT}} = \boldsymbol{\sigma}^S = (0.4, 1.0, 1.0, 0.3, 0.3)'$. The EM-UV (equal mean and unequal variance) scenario had $\boldsymbol{\mu}^{\text{RCT}} = \boldsymbol{\mu}^S = (-2, -3, -2.5, -2, -1.5)'$, $\boldsymbol{\sigma}^{\text{RCT}} = (0.4, 1.0, 1.0, 0.3, 0.3)'$, and $\boldsymbol{\sigma}^S = (1.2, 0.5, 0.5, 0.9, 0.9)'$. The UM-UV (unequal mean and unequal variance) scenario had $\boldsymbol{\mu}^{\text{RCT}} = (-2, -3, -2.5, -2, -1.5)'$, $\boldsymbol{\mu}^S = (-1, -2.5, -2, -2.5, -2)'$, $\boldsymbol{\sigma}^{\text{RCT}} = (0.4, 1.0, 1.0, 0.3, 0.3)'$, and $\boldsymbol{\sigma}^S = (1.2, 0.5, 0.5, 0.9, 0.9)'$.

Once the complete dataset of RCTs was generated, we excluded treatment arms to create a realistic (partially missing) NMA dataset as illustrated in Figure 5, with two types of missingness: missing completely at random (MCAR) and 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 5.

We used the prior specifications in Table 1 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 (σ_t), and log odds ratio between the i^{th} and j^{th} treatments (LOR_{ij}). To measure the methods' performance, we followed the instructions provided by Morris, White and Crowther (2019). In particular, we used bias, mean squared error (MSE), and the 95% CrI's coverage probability (CP) and length (CrIL). Detailed simulation results for the different performance measures and corresponding Monte Carlo standard errors are in Appendix C, as well as the effective sample size, which describes the amount of information in the MCMC samples.

5.2. Simulation results.

Table 3 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, we summarized overall measures across all treatments or across all pairs of comparisons for each of bias, MSE, and CP. 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 D1 in Appendix D presents the same summaries of simulation results under MCAR. Figure D1 in Appendix D 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, σ_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 3), 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 trials 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 trials was not reliable (EM-UV and UM-UV). Similar performance patterns were present under the MCAR mechanism (Table D1 in Appendix D), 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 D1 in Appendix D). 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 trials, 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 trials differed from true variances in the RCTs, e.g., the elementwise ratios were $\boldsymbol{\sigma}^{\text{RCT}}/\boldsymbol{\sigma}^{\text{S}} = (1/3, 2, 2, 1/3, 1/3)'$. The performance of the other three borrowing strategies depended largely on the reliability of the single-arm trials.

6. Summary and discussion.

This paper has proposed and discussed different strategies to incorporate single-arm trials into AB-NMA to mitigate the prevalent "lack of information" problem. We have performed extensive simulation studies to explore whether it is preferable to choose a full borrowing

strategy or one of the adaptive borrowing methods (commensurate prior), and whether to borrow mean information, variance information, or both. The simulation studies considered scenarios when information from single-arm trials 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 improve efficiency even in the presence of modestly discordant variance information, by facilitating partial pooling of variance information from single-arm trials, 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. Although we did not examine FBM in our paper, it could be expected that FBM would perform even worse than CPM if mean information from single-arm trials was biased. To the best of our knowledge, this is the first proposal in 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 trials.

Some researchers have reservations about the AB-NMA approach because it includes random study intercepts, which may lead to bias when differences exist between trials with different designs (White et al., 2019). Hence, we compared our proposed approach (CPV) with the classic CB-NMA model and the CB-NMA with random study-specific baseline intercepts (CB-2) in the case study. Appendix F presents the results. Overall, the point estimates of LOR and p_t from the CPV method were consistent with those from the classic CB and CB-2 models, except for estimands related to treatments 6 (ICI+ICC) and 7 (2ICIs). This was expected because of substantial differences in the ICI drug or the dose of the drug used in clinical trials that compared ICI+ICC or 2ICIs with other treatments, which led to large uncertainties in safety profile. On the other hand, by including additional information from single-arm trials, the CPV method delivered overall smaller CrILs for LORs and p_ts compared with the classic CB and CB-2 methods. In summary, our proposed method was compatible with the CB-NMA approaches on point estimation and achieved more certainty by borrowing information, at least in this case study.

We have focused on commensurate priors to synthesize RCTs and single-arm trials 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 (Murray, Hobbs and Carlin, 2015). We did a sensitivity analysis to assess how much importance should be placed on the single-arm trials (see Appendix E) and found that any value between 0.1 to 0.5 would be suitable for p^{v} in the CPV method, while the choice of p^{m} is quite sensitive in the CPM method. Future research is needed specifically for different combinations of (p^{v}, R^{v}, $s_{\mathrm{l}}^{\mathrm{v}}$, $s_{\mathrm{u}}^{\mathrm{v}}$) for AB-NMA. Also, existing methods cannot assess the importance of single-arm trials in 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 (Zhang et al., 2017b), 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 trials 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 trials into AB-NMA. For example, the extrapolation strategy for meta-analyses proposed by Röver, Wandel and Friede (2018) 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, Wandel and Friede (2018), it is possible to borrow only variance information by averaging just two models: NB and FBV. Another method is power priors (Ibrahim et al., 2015; Zhang et al., 2019) that incorporate supplemental information by raising the likelihood of the single-arm trials to a power $a \in [0, 1]$. 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. Zhang et al. (2019) used simulation studies to compare the hierarchical power prior (HPP) and hierarchical commensurate prior (HCP) in meta-analysis of binary data. They concluded that the performance of HCP was better in terms of relative bias and stable throughout all their simulation scenarios. Also, by building on the design-by-treatment interaction model (White et al., 2012; Jackson et al., 2014), we could treat the single-arm trials as inconsistent with the RCTs, and set up the following random inconsistency effects model:

$$\begin{aligned} \text{logit}(p_{kt}) &= \mu_t + \beta_{kt}, k = 1, \ldots, K; \\ \text{logit}\left(p^{S}_{jt}\right) &= \mu_t + \beta^{S}_{jt} + \omega_t, j = 1, \ldots, J; \end{aligned} \tag{8}$$

where μ_t is the overall logit event probability for each treatment, ω_t is a design-by-treatment interaction effect reflecting inconsistency between single-arm trials and RCTs, and β_{kt} and β^{S}_{jt} are trial-by-treatment interaction effects to reflect between-trial heterogeneity. The ω_t's could be treated as fixed effects or random effects. In the random-effects approach, these $\omega_t \{t = 1, \ldots, T\}$ could be independent or dependent with correlations representing the similarity across different treatments of inconsistency between single-arm trials and RCTs.

Fourth, Turner et al. (2019) proposed to incorporate external evidence in CB-NMA by using informative priors specified based on previously published evidence describing between-trial heterogeneity. A similar idea could be used in AB-NMA by first obtaining a posterior distribution for σ^{S}_t from single-arm trials, then using this as an informative prior for σ_t. However, small B^{S}_t or B_t may still provide little improvement in estimation. Another widely used approach in CB-NMA to borrow information from single-arm trials or observational studies is aggregate-level covariate matching (ALCM), which was thoroughly discussed in Leahy et al. (2019). In short, covariate information is first used to match single-arm trials to form constructed RCTs; when analyzing the existing network of RCTs, these constructed RCTs are also incorporated by various statistical methods. One advantage of this method over our approach is that it can fill the gap between two disconnected evidence networks (Schmitz et al., 2018), while CPV cannot. Currently, the CB-NMA approach is

also more widely accepted by health technology assessment agencies than the AB-NMA approach. However, the CPV method only incorporates variance information of absolute effects from single-arm trials; in contrast, the ALCM method borrows both mean and variance information of constructed contrast effects, which may cause potentially larger bias. Also, incorporating variance information from single-arm trials in the CPV method, even though this method may be subject to potential bias of variance information from the perspective of randomized clinical trial designs, could help us understand the variability of absolute treatment effects in the real world applications. Results based on CPV method may serve as a supplementary source of evidence to make better decisions. Moreover, we could also use covariate information to match single-arm trials and treatment arms from RCTs to determine how much information we should borrow in the CPV method.

Finally, we could also develop power priors, commensurate priors, or Bayesian model averaging methods to incorporate supplemental information under the framework of the CB-NMA. The triangle inequalities on between-trial standard deviations could make prior specifications more complicated (Lu and Ades, 2009).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

This research was supported in part by the U.S. National Institutes of Health grant R01 LM012982 and the Clinical and Translational Science Award UL1TR002494. We thank Professor Leonhard Held, Professor Georgia Salanti, an Associate Editor, and an anonymous referee for many helpful comments that substantially improved the quality of this article.

REFERENCES

Barnard J, McCulloch R and Meng XL (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica 10 1281–1311.

Begg CB and Pilote L (1991). A model for incorporating historical controls into a meta-analysis. Biometrics 47 899–906. [PubMed: 1742445]

Chen M-H and Ibrahim JG (2000). Power prior distributions for regression models. Statistical Science 15 46–60.

Chen M-H, Ibrahim JG, Lam P, Yu A and Zhang Y (2011). Bayesian design of noninferiority trials for medical devices using historical data. Biometrics 67 1163–1170. [PubMed: 21361889]

de Valpine P (2016). Comparisons between NIMBLE, JAGS and Stan for the election88 example (“full” version) from the book known as “Applied Regression Modeling” (Gelman and Hill 2007). https://nature.berkeley.edu/~pdevalpine/MCMC_comparisons/some_ARM_comparisons/election/nimble_election88_comparisons.html. Accessed: 2020-12-18.

de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT and Bodik R (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. Journal of Computational and Graphical Statistics 26 403–413.

Dias S and Ades AE (2016). Absolute or relative effects? Arm-based synthesis of trial data. Research Synthesis Methods 7 23–28. [PubMed: 26461457]

Dias S, Sutton AJ, Ades AE and Welton NJ (2013). Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical Decision Making 33 607–617. [PubMed: 23104435]

Duan Y, Ye K and Smith EP (2005). Evaluating water quality using power priors to incorporate historical information. Environmetrics 17 95–106.

Efthimiou O, Mavridis D, Debray TPA, Samara M, Belger M, Siontis GCM, Leucht S and and GS (2017). Combining randomized and non-randomized evidence in network meta-analysis. Statistics in Medicine 36 1210–1226. [PubMed: 28083901]

Egger M, Davey Smith G and Altman DG, eds. (2001). Systematic reviews in health care. BMJ Publishing Group.

Gamalo MA, Tiwari RC and LaVange LM (2013). Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products. Pharmaceutical Statistics 13 25–40. [PubMed: 23913880]

Gamalo-Siebers M, Savic J, Basu C, Zhao X, Gopalakrishnan M, Gao A, Song G, Baygani S, Thompson L, Xia HA, Price K, Tiwari R and Carlin BP (2017). Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. Pharmaceutical Statistics 16 232–249. [PubMed: 28448684]

Geisser S and Eddy WF (1979). A predictive approach to model selection. Journal of the American Statistical Association 74 153–160.

Hanson TE, Branscum AJ and Johnson WO (2011). Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. Lifetime Data Analysis 17 3–28. [PubMed: 20369294]

Hobbs BP, Sargent DJ and Carlin BP (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. Bayesian Analysis 7 639–674. [PubMed: 24795786]

Hobbs BP, Carlin BP, Mandrekar SJ and Sargent DJ (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics 67 1047–1056. [PubMed: 21361892]

Hong H, Fu H and Carlin BP (2018). 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 1047–1069.

Hong H, Chu H, Zhang J and Carlin BP (2016a). A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. Research Synthesis Methods 7 6–22. [PubMed: 26536149]

Hong H, Chu H, Zhang J and Carlin BP (2016b). 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 29–33. [PubMed: 26461816]

Hueber W, Sands BE, Lewitzky S, Vandemeulebroecke M, Reinisch W, Higgins PDR, Wehkamp J, Feagan BG, Yao MD, Karczewski M, Karczewski J, Pezous N, Bek S, Bruin G, Mellgard B, Berger C, Londei M, Bertolino AP, Tougas G and and SPLT (2012). 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 1693–1700. [PubMed: 22595313]

Ibrahim JG, Chen M-H, Gwon Y and Chen F (2015). The power prior: theory and applications. Statistics in Medicine 34 3724–3749. [PubMed: 26346180]

Jackson D, Barrett JK, Rice S, White IR and Higgins JPT (2014). A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. Statistics in Medicine 33 3639–3654. [PubMed: 24777711]

Jaff MR, Nelson T, Ferko N, Martinson M, Anderson LH and Hollmann S (2017). Endovascular interventions for femoropopliteal peripheral artery disease: a network meta-analysis of current technologies. Journal of Vascular and Interventional Radiology 28 1617–1627. [PubMed: 29031986]

Johnson DB, Chandra S and Sosman JA (2018). Immune checkpoint inhibitor toxicity in 2018. JAMA 320 1702–1703. [PubMed: 30286224]

Kaizer AM, Hobbs BP and Koopmeiners JS (2018). A multi-source adaptive platform design for testing sequential combinatorial therapeutic strategies. Biometrics 74 1082–1094. [PubMed: 29359450]

Kaizer AM, Koopmeiners JS and Hobbs BP (2018). Bayesian hierarchical modeling based on multisource exchangeability. Biostatistics 19 169–184. [PubMed: 29036300]

Kontopantelis E, Springate DA and Reeves D (2013). A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. PLoS ONE 8 e69930. [PubMed: 23922860]

Leahy J, Thom H, Jansen JP, Gray E, O'Leary A, White A and Walsh C (2019). Incorporating single-arm evidence into a network meta-analysis using aggregate level matching: assessing the impact. Statistics in Medicine 38 2505–2523. [PubMed: 30895655]

Li Z and Begg CB (1994). Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. Journal of the American Statistical Association 89 1523–1527.

Lin L, Chu H and Hodges JS (2016). Sensitivity to excluding treatments in network meta-analysis. Epidemiology 27 562–569. [PubMed: 27007642]

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

Lu G and Ades AE (2004). Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine 23 3105–3124. [PubMed: 15449338]

Lu G and Ades AE (2006). Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association 101 447–459.

Lu G and Ades AE (2009). Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics 10 792–805. [PubMed: 19687150]

Lunn D, Jackson C, Best N, Spiegelhalter D and Thomas A (2010). The BUGS book. Taylor & Francis Inc.

Mathes T and Kuss O (2018). A comparison of methods for meta-analysis of a small number of studies with binary outcomes. Research Synthesis Methods 9 366–381. [PubMed: 29573180]

Morris TP, White IR and Crowther MJ (2019). Using simulation studies to evaluate statistical methods. Statistics in Medicine 38 2074–2102. [PubMed: 30652356]

Murray TA, Hobbs BP and Carlin BP (2015). Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. The Annals of Applied Statistics 9 1549–1570. [PubMed: 26557211]

Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH and Salanti G (2014). Characteristics of networks of interventions: a description of a database of 186 published networks. PLoS ONE 9 e86754. [PubMed: 24466222]

Phillippo DM, Dias S, Ades AE and Welton NJ (2020). Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study. Statistics in Medicine 39 4885–4911. [PubMed: 33015906]

Röver C, Wandel S and Friede T (2018). Model averaging for robust extrapolation in evidence synthesis. Statistics in Medicine 38 674–694. [PubMed: 30302781]

Salanti G, Ades AE and Ioannidis JPA (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. Journal of Clinical Epidemiology 64 163–171. [PubMed: 20688472]

Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D and Neuenschwander B (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics 70 1023–1032. [PubMed: 25355546]

Schmitz S, Maguire Á, Morris J, Ruggeri K, Haller E, Kuhn I, Leahy J, Homer N, Khan A, Bowden J, Buchanan V, O'Dwyer M, Cook G and Walsh C (2018). The use of single armed observational data to closing the gap in otherwise disconnected evidence networks: a network meta-analysis in multiple myeloma. BMC Medical Research Methodology 18 66. [PubMed: 29954322]

Spiegelhalter DJ, Best NG, Carlin BP and van der Linde A (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 583–639.

Thom HH, Capkun G, Cerulli A, Nixon RM and Howard LS (2015). Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an

application to pulmonary arterial hypertension. BMC Medical Research Methodology 15 34. [PubMed: 25887646]

Turner RM, Domínguez-Islas CP, Jackson D, Rhodes KM and White IR (2019). Incorporating external evidence on between-trial heterogeneity in network meta-analysis. Statistics in Medicine 38 1321–1335. [PubMed: 30488475]

Wang Z, Lin L, Hodges JS and Chu H (2020a). The impact of covariance priors on arm-based Bayesian network meta-analyses with binary outcomes. Statistics in Medicine 39 2883–2900. [PubMed: 32495349]

Wang Z, Lin L, Zhao S and Chu H (2020b). Nmaplateplot: The Plate Plot for Network Meta-Analysis Results R package version 1.0.0.

Wang Z, Lin L, Hodges JS, MacLehose R and Chu H (2021). A variance shrinkage method improves arm-based Bayesian network meta-analysis. Statistical Methods in Medical Research 30 151–165. [PubMed: 32757707]

Welton NJ, Sutton AJ, Cooper NJ and Abrams KR (2012). Evidence synthesis for decision making in healthcare. Wiley-Blackwell.

White IR, Barrett JK, Jackson D and Higgins JPT (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Research Synthesis Methods 3 111–125. [PubMed: 26062085]

White IR, Turner RM, Karahalios A and Salanti G (2019). A comparison of arm-based and contrast-based models for network meta-analysis. Statistics in Medicine 38 5197–5213. [PubMed: 31583750]

Xu C, Chen Y-P, Du X-J, Liu J-Q, Huang C-L, Chen L, Zhou G-Q, Li W-F, Mao Y-P, Hsu C, Liu Q, Lin A-H, Tang L-L, Sun Y and Ma J (2018). Comparative safety of immune checkpoint inhibitors in cancer: systematic review and network meta-analysis. BMJ k4226. [PubMed: 30409774]

Zeger SL, Liang K-Y and Albert PS (1988). Models for longitudinal data: a generalized estimating equation approach. Biometrics 44 1049–1060. [PubMed: 3233245]

Zhang J, Carlin BP, Neaton JD, Soon GG, Nie L, Kane R, Virnig BA and Chu H (2014). Network meta-analysis of randomized clinical trials: reporting the proper summaries. Clinical Trials 11 246–262. [PubMed: 24096635]

Zhang J, Chu H, Hong H, Virnig BA and Carlin BP (2017a). Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness. Statistical Methods in Medical Research 26 2227–2243. [PubMed: 26220535]

Zhang D, Chen M-H, Ibrahim JG, Boye ME and Shen W (2017b). 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. [PubMed: 28239247]

Zhang J, Ko C-W, Nie L, Chen Y and Tiwari R (2019). Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. Statistical Methods in Medical Research 28 1293–1310. [PubMed: 29433407]

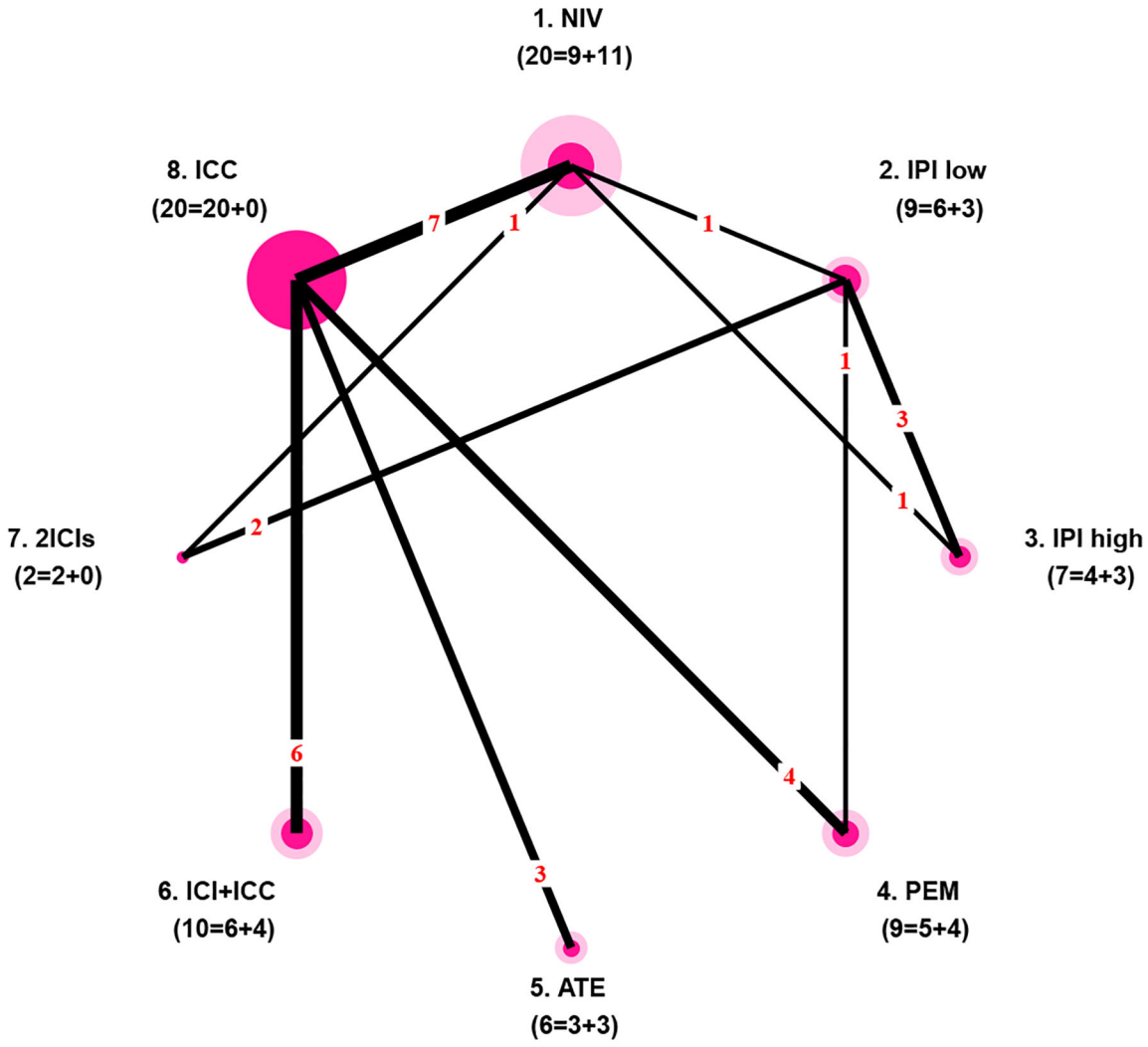


Fig 1.

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. Node radius is proportional to the number of RCTs including the regimen (dark inner circle) plus the number of single-arm trials of the regimen (light outer circle). Edge thickness is proportional to the number of direct comparisons. Numbers in parentheses under a regimen name are the number of RCTs and the number of single-arm trials including the regimen. For example, 9 RCTs and 11 single-arm trials investigate nivolumab.

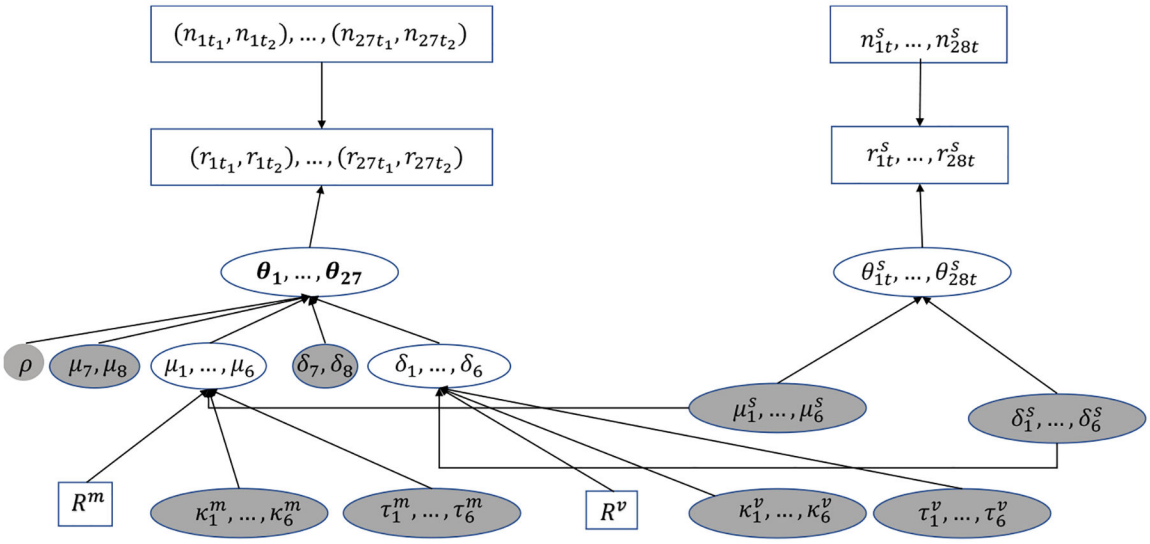
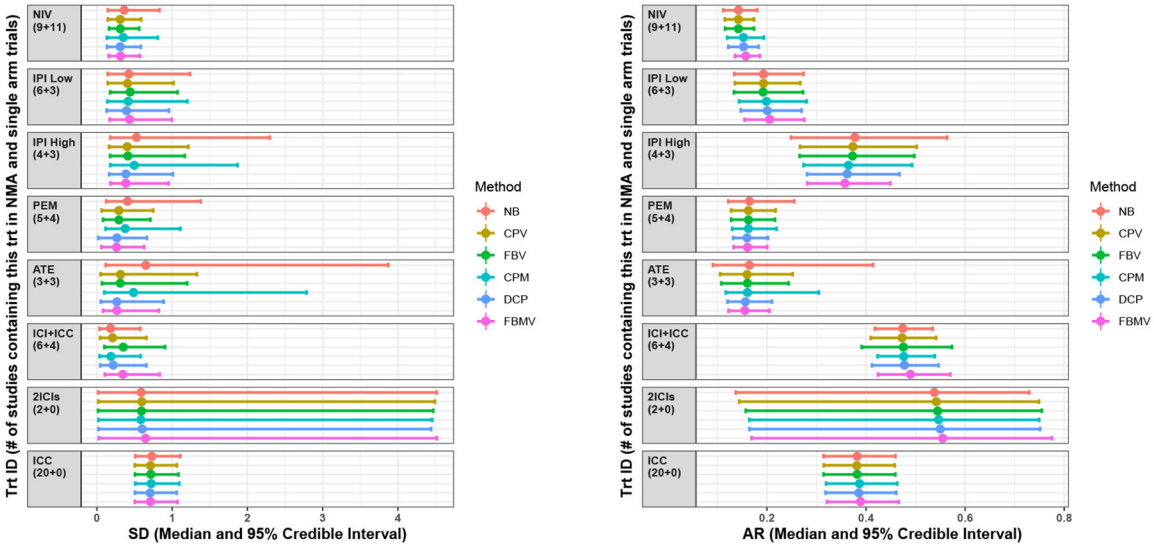


Fig 2.

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.



(a) standard deviation σ_t (b) absolute risk p_t

Fig 3.

Results for the dataset of the safety of ICIs in cancer treatment: forest plot of posterior estimates of standard deviations σ_t and absolute risks p_t (posterior medians with 95% credible intervals). Different colors indicate different methods. The y-axis represents regimen abbreviations, with the number of RCTs (B_t) and single-arm trials (B_t^s) in parentheses.

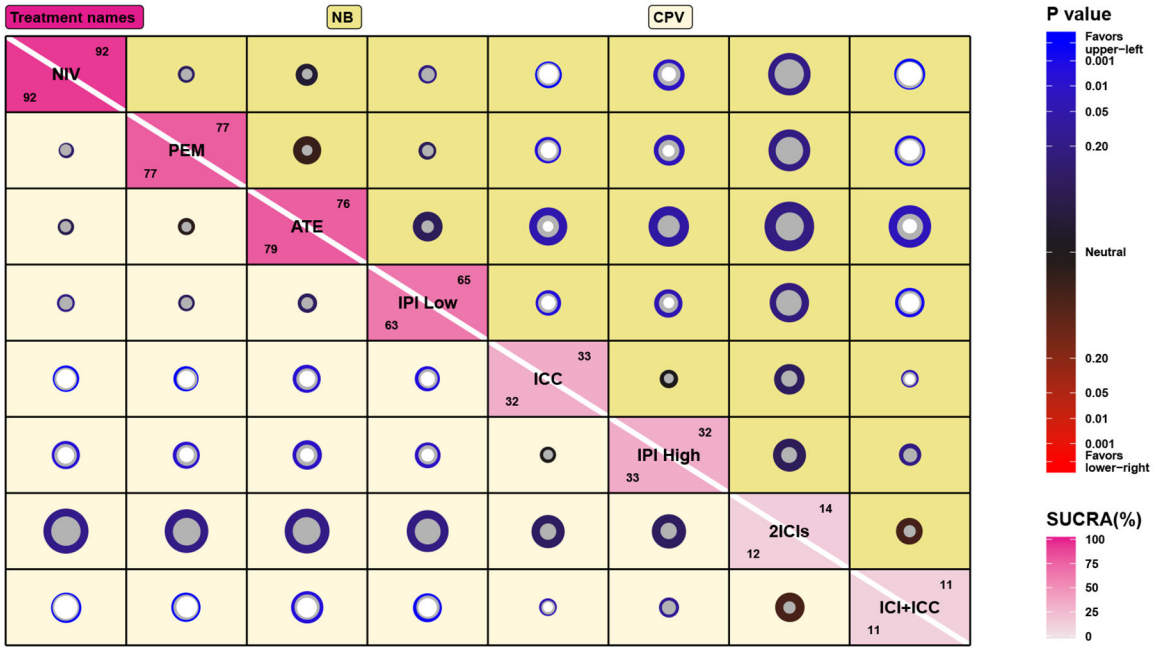


Fig 4.

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.

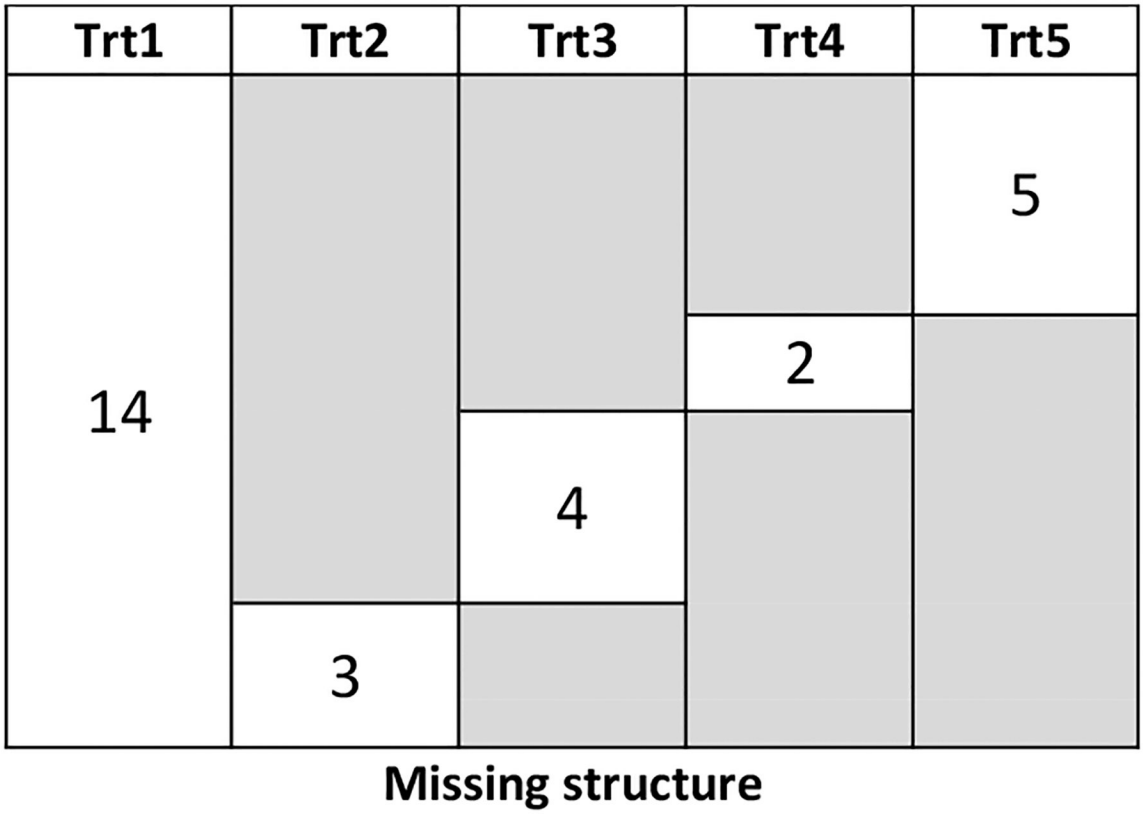


Fig 5.

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 1

Summary of prior specifications and assumptions for μ_t, σ_t, μ_t^s, σ_t^s, and the correlation matrix **P** in six different models.

Model	Parameter				
	$\boldsymbol{\mu_t, t = 1,\ldots,T}$	$\boldsymbol{\sigma_t, t = 1,\ldots,T}$	$\mu_t^s, t = 1,\ldots,T$	$\sigma_t^s, t = 1,\ldots,T$	$\mathbf{P} = \{\boldsymbol{\rho_{ij}}\}$
NB	$\mu_t \sim N(0,100^2)$	$\sigma_t \sim U(0,5)$	NA	NA	$\rho_{ij} = \rho\ (i \neq j)$ and $\rho \sim U\left(-\frac{1}{T-1}, 1\right)$ for all models
FBMV	$\mu_t \sim N(0,100^2)$	$\sigma_t \sim U(0,5)$	$\mu_t = \mu_t^s$	$\sigma_t = \sigma_t^s$	
FBV	$\mu_t \sim N(0,100^2)$	$\sigma_t \sim U(0,5)$	$\mu_t^s \sim N\left(0, 100^2\right)$	$\sigma_t = \sigma_t^s$	
CPM	Equation (4) with $p^m = 0.5$, $R^m = 2500$, $s_1^m = 0$, and $s_u^m = 2$	$\sigma_t \sim U(0,5)$	$\mu_t^s \sim N\left(0, 100^2\right)$	$\sigma_t^s \sim U(0,5)$	
CPV	$\mu_t \sim N(0,100^2)$	Equation (6) with $p^v = 0.5$, $R^v = 2500$, $s_1^v = 0$, and $s_u^v = 2$	$\mu_t^s \sim N\left(0, 100^2\right)$	$\sigma_t^s \sim U(0,5)$	
DCP	Equation (4) with $p^m = 0.5$, $R^m = 2500$, $s_1^m = 0$, and $s_u^m = 2$	Equation (6) with $p^v = 0.5$, $R^v = 2500$, $s_1^v = 0$, and $s_u^v = 2$	$\mu_t^s \sim N\left(0, 100^2\right)$	$\sigma_t^s \sim U(0,5)$	

Table 2

Analysis of ICI safety in cancer treatment: posterior medians and 95% credible intervals by 6 different models (NB, CPV, FBV, CPM, DCP, and FBMV). The estimated parameters include 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 (σ_t), and selected log odds ratios LOR_{ij} comparing treatments i and j. 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)
σ_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)
σ_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)
σ_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)
σ_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)
σ_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)
σ_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)
σ_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)
σ_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)

Parameter	Posterior median (95% credible interval)					
	NB	CPV	FBV	CPM	DCP	FBMV
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

Ann Appl Stat. Author manuscript; available in PMC 2022 August 26.

Table 3.

Simulation results comparing data generated under four different scenarios (EM-EV, UM-EV, EM-UV, and UM-UV) with MAR missingness of treatment arms. 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). The value in the column for bias and the row for LOR_{ij} is calculated as $\sum_{i\neq j}|\text{bias}(LOR_{ij})|$; the value in the column for coverage probability and in the row for LOR_{ij} is calculated as $\sum_{i\neq j}(0.95-\text{CP}(LOR_{ij}))_{+}$.; the value in the column for coverage probability and in the row for LOR_{25} is $\text{CP}(LOR_{25})$.

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
σ_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
σ_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
σ_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
σ_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
σ_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

Parameter	Truth	Bias					Mean squared error					Coverage probability				
		NB	CPV	FBV	CPM	DCP	NB	CPV	FBV	CPM	DCP	NB	CPV	FBV	CPM	DCP
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
σ_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
σ_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
σ_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