RESEARCH METHODS

Prior Choices of Between-Study Heterogeneity in Contemporary Bayesian Network Meta-analyses: an Empirical Study

Kristine J. Rosenberger, MS[1], Aiwen Xing, MS[1], Mohammad Hassan Murad, MD[2], Haitao Chu, MD, PhD[3], and Lifeng Lin, PhD[1] (iD)

[1]Department of Statistics, Florida State University, Tallahassee, FL, USA; [2]Evidence-Based Practice Center, Mayo Clinic, Rochester, MN, USA; [3]Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA.

BACKGROUND: Network meta-analysis (NMA) is a popular tool to compare multiple treatments in medical research. It is frequently implemented via Bayesian methods. The prior choice of between-study heterogeneity is critical in Bayesian NMAs. This study evaluates the impact of different priors for heterogeneity on NMA results.

METHODS: We identified all NMAs with binary outcomes published in The BMJ, JAMA, and The Lancet during 2010–2018, and extracted information about their prior choices for heterogeneity. Our primary analyses focused on those with publicly available full data. We re-analyzed the NMAs using 3 commonly-used non-informative priors and empirical informative log-normal priors. We obtained the posterior median odds ratios and 95% credible intervals of all comparisons, assessed the correlation among different priors, and used Bland–Altman plots to evaluate their agreement. The kappa statistic was also used to evaluate the agreement among these priors regarding statistical significance.

RESULTS: Among the selected Bayesian NMAs, 52.3% did not specify the prior choice for heterogeneity, and 84.1% did not provide rationales. We re-analyzed 19 NMAs with full data available, involving 894 studies, 173 treatments, and 395,429 patients. The correlation among posterior median (log) odds ratios using different priors were generally very strong for NMAs with over 20 studies. The informative priors produced substantially narrower credible intervals than non-informative priors, especially for NMAs with few studies. Bland–Altman plots and kappa statistics indicated strong overall agreement, but this was not always the case for a specific NMA.

CONCLUSIONS: Priors should be routinely reported in Bayesian NMAs. Sensitivity analyses are recommended to examine the impact of priors, especially for NMAs with relatively small sample sizes. Informative priors may produce substantially narrower credible intervals for such NMAs.

KEY WORDS: Bayesian analysis; heterogeneity; network meta-analysis; prior distribution; sensitivity analysis.

INTRODUCTION

Network meta-analysis (NMA) is a statistical method often used to draw conclusions about multiple-treatment comparisons.[1–3] It simultaneously synthesizes both direct and indirect evidence, where the direct evidence comes from head-to-head trials while the indirect evidence comes from indirect comparisons with common comparators.[4,5] For example, the comparison between two active drugs A and B can be informed from indirect comparisons of A vs. C and B vs. C, where C may be placebo or standard care, or from direct comparison in clinical trials comparing A vs. B.

In addition to the advantage of combining direct and indirect evidence, NMAs improve the precision of estimates (i.e., make the confidence/credible intervals narrower).[6,7] This precision however is affected by the amount of heterogeneity between studies, because heterogeneity is modeled into the uncertainty and impacts the width of confidence/credible intervals.[8,9] Currently, many NMAs are performed via a Bayesian framework that uses a prior distribution for the between-study heterogeneity.[10] Some NMAs use the traditionally non- or weakly informative priors for heterogeneity.[11–13] Recently, Turner et al.[14] have suggested informative log-normal priors based on a large database of conventional pairwise meta-analyses in the Cochrane Library. These empirical priors have the potential to improve the precision of the treatment effect estimates, especially when the number of studies is small.

The choice of prior distribution is important and should be explicitly reported according to the PRISMA-NMA statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Network Meta-Analysis).[15] It is critical that authors of NMAs describe the details of choices and assumptions made to select the prior distribution for transparency purposes and to allow reproducibility of the work. Nevertheless, it has been found that the rapid growth of NMAs published in recent years was not accompanied with better methodological and reporting quality.[16,17] Therefore, we

conducted this empirical study to assess recent NMAs published in high-impact medical journals for the quality of reporting heterogeneity priors (in terms of distribution and rationale) and to evaluate how NMAs' conclusions would differ based on applying various commonly used prior distributions.

METHODS

Data Collection

We designed and executed a literature search in July, 2019, for research articles published in *The BMJ*, *JAMA*, and *The Lancet* between January 1, 2010, and December 31, 2018, using the terms "network meta-analysis," "network meta-analyses," "multiple-treatment comparison," "multiple-treatment meta-analysis," and "multiple-treatment meta-analyses." If Bayesian models were used, we examined if the original articles gave information about prior choices for heterogeneity and their rationales.

In our primary analyses, we excluded methodological reviews that did not present original data. The outcome type was restricted to be binary, because the heterogeneity may depend on the measure scale for other outcome types (e.g., for continuous outcomes) and needed to be modeled on a case-by-case basis. In addition, we focused on articles whose full NMA datasets were publicly available. Originally reported effect measures, outcomes, studies, treatments, event counts, sample sizes, statistical methods (frequentist or Bayesian) used for NMAs, and prior distributions (for Bayesian NMAs) were obtained from the published articles (and the corresponding supplemental files). If the authors used the Bayesian method but did not report their prior distributions, we contacted them for information about the priors. An article may report multiple NMAs with various outcomes; our analyses focused on the primary outcome. If no primary outcome was specified, we used the NMA with the largest number of studies. In addition, regardless the original effect measures, we re-analyzed the collected data based on the odds ratio (OR) (on a logarithmic scale) for a consistent comparison across NMAs.

Prior Distribution Choices for Heterogeneity

The Bayesian framework treats unknown parameters, such as overall treatment effects and heterogeneity variances, as random variables and attempts to estimate them via the assignment of prior distributions. This is commonly implemented via the Markov chain Monte Carlo (MCMC) algorithm.[18,19] The Supplementary Material presents the Bayesian contrast-based random-effects model to perform an NMA (Appendix A). Among the various parameters to be estimated, the heterogeneity variance, denoted by τ^2, plays a critical role, because researchers often have diverse opinions toward its prior choice and it sometimes has influential impact on credible intervals (CrIs) of

Table 1 Summary of Prior Distributions for the Heterogeneity Component (Variance τ^2 or Standard Deviation τ) for Odds Ratios

Prior distribution	Used for	Hyper-parameter
Inverse-gamma, $IG(\alpha, \beta)$	τ^2	$\alpha = \beta = 0.1$; $\alpha = \beta = 0.01$; or $\alpha = \beta = 0.001$.
Uniform, $U(0, c)$	τ	$c = 2$; $c = 5$; or $c = 10$.
Half-normal, $HN(0, \sigma^2)$	τ	$\sigma^2 = 0.5$; $\sigma^2 = 1$; or $\sigma^2 = 2$.
Informative log-normal, $LN(\mu, \sigma^2)$	τ^2	Pharmacological vs. placebo/control comparison: $\mu = -4.06$, $\sigma = 1.45$ (all-cause mortality); $\mu = -3.02$, $\sigma = 1.85$ (semi-objective outcome); $\mu = -2.13$, $\sigma = 1.58$ (subjective outcome). Pharmacological vs. pharmacological comparison: $\mu = -4.27$, $\sigma = 1.48$ (all-cause mortality); $\mu = -3.23$, $\sigma = 1.88$ (semi-objective outcome); $\mu = -2.34$, $\sigma = 1.62$ (subjective outcome). Non-pharmacological comparison: $\mu = -3.93$, $\sigma = 1.51$ (all-cause mortality); $\mu = -2.89$, $\sigma = 1.91$ (semi-objective outcome); $\mu = -2.01$, $\sigma = 1.64$ (subjective outcome).

treatment comparisons. On the other hand, researchers generally have consensus about the prior for treatment effects (log ORs in this study), which is usually non-informative and follows a normal distribution with mean 0 and a very large variance (e.g., 100^2).

To re-analyze the collected datasets, we considered three different non-informative (or arguably, weakly informative) prior distributions for the heterogeneity variance or standard deviation: the inverse-gamma, uniform, and half-normal distributions. Table 1 summarizes the multiple choices.

The inverse-gamma prior $IG(\alpha, \beta)$ is conjugate (that is, it produces a posterior also in the inverse-gamma family) and therefore may facilitate the computation in the MCMC algorithm. It also has the potential to improve both stability and convergence, and may be useful for sparse data.[20] The hyper-parameters for both α and β (determining distribution shape and scale, respectively) are conventionally assigned to some value close to 0. As both hyper-parameters approach 0, it leads to a flat distribution for the heterogeneity variance on a logarithmic scale. We consider three choices of the hyper-parameters, i.e., 0.1, 0.01, and 0.001.

The uniform prior $U(0, c)$ is another commonly used prior for the heterogeneity standard deviation τ. Here, c denotes the upper bound of the uniform distribution, and is assigned to values 2, 5, and 10 in our analyses; these are common choices for log ORs in practice. We also considered the half-normal prior $HN(0, \sigma^2)$ for τ.[21] This distribution is generated by taking

the absolute value of a random variable that follows the normal distribution $N(0, \sigma^2)$. The hyper-parameter σ^2 determines the range of heterogeneity,[22] and is assigned values 0.5, 1, and 2 in our analyses.

In addition to the non-informative priors above for heterogeneity, we considered the empirical informative priors derived by Turner et al.[14] for log ORs, which were grouped based on outcome type and treatment comparison type. Specifically, the treatment comparisons were classified into three groups, i.e., pharmacological treatment vs. placebo/control, pharmacological treatment vs. pharmacological treatment, and comparisons involved with non-pharmacological treatments (e.g., medical devices, surgical procedures). The outcome types were classified as all-cause mortality, semi-objective outcomes (e.g., cause-specific mortality, major morbidity events, obstetric outcomes), and subjective outcomes (e.g., pain, mental health outcomes, general physical health). For NMAs containing multiple comparison types, the type of pharmacological vs. pharmacological treatment comparison was used for the primary analyses.

Statistical Analyses

For each dataset, we re-performed the random-effects NMAs with the non-informative priors and the informative priors in Table 1 for the heterogeneity parameter τ or τ^2. The non-informative prior $N(0, 100^2)$ was used for all treatment effects (log ORs), and we assumed consistency between direct and indirect evidence in each NMA. In addition, all treatment comparisons in each NMA were assumed to share a common heterogeneity variance.[5] Of note, the NMA model used in this article was contrast-based, because it is currently the most widely used model and the informative priors were derived under the contrast-based framework. Many alternatives, such as the arm-based model (which focuses on estimating each treatment arm's absolute effect), may be also used for NMAs.[23]

The Bayesian analyses were conducted with R (version 3.6.2) package "rjags" (version 4-9). The models were implemented via the MCMC algorithm with three chains[24–27]; each chain contained a 50,000-run burn-in period for achieving stabilization and convergence. The samples generated during the burn-in period were discarded prior to the final analyses; the final posterior distributions for each NMA were based on a run of 200,000 updates after the burn-in period. We checked trace plots for assessing MCMC convergence. Trace plots with certain long-term trends or drifts, instead of stable up-and-down variation, may indicate non-convergence; see Figs. S1–S4 in the Supplementary Material for illustrations. The MCMC may not converge well in cases such as extreme posterior samples of ORs (produced seemingly due to many zero even counts) or improper priors. When Markov chains converged well, the posterior medians and 95% equal-tailed CrIs can be reliably used as estimates of the parameters of interest. CrIs of log ORs not covering 0 indicated significant

treatment comparisons. We obtained the posterior estimates for all treatment comparisons and the heterogeneity variance in each NMA. We also calculated the width of 95% CrI of log OR for each comparison, which implied the estimate's precision.

Correlation coefficients between the non-informative priors and informative priors were calculated for both point estimates (posterior median log ORs) and CrI widths for each NMA and for all NMAs combined. Bland–Altman plots were used to evaluate the agreement between these results. The kappa statistic, κ, was also calculated to quantify the agreement of statistical significance between the treatment effects produced by the different priors. This statistic is upper bounded by 1; roughly, $\kappa < 0$ indicates no agreement, and κ within 0–0.4, 0.4–0.6, and 0.6–1 indicates weak, moderate, and strong agreement, respectively.[28,29]

Secondary analyses were performed for NMAs which contained a placebo or control treatment. Among these NMAs, we additionally considered the informative prior of the comparison type of pharmacological treatments vs. placebo/control (Table 1).

RESULTS

Basic Characteristics

The literature search identified 67 research articles containing NMAs. Of the 44 NMAs that used the Bayesian framework, 52.3% of the NMAs did not explicitly provide the prior distributions of heterogeneity, and 84.1% did not provide rationales for the prior choices (Table S1 in the Supplementary Material). A total of 19 NMAs met inclusion criteria for our primary analyses (Fig. 1). Total sample sizes in the selected NMAs ranged from 792 to 111,282; numbers of treatments ranged from 3 to 23; and numbers of studies ranged from 7 to 473. We denoted each NMA by the first author's surname with the publication year of the corresponding article. Table 2 presents summaries of these NMAs; the complete references of these NMAs are in the Supplementary Material (Appendix B). Of note, the NMA of Wu 2013 contained zero events in many treatment arms, causing poor MCMC convergence in our re-analyses; thus, we only present the results of the remaining 18 NMAs in the following.

Overall Impact of Priors

Figure 2 compares the posterior median ORs and 95% CrI widths produced by non-informative priors with those by informative priors among all 18 NMAs. There was a nearly perfect correlation between posterior median (log) ORs by each type of non-informative priors and those by the informative priors; the correlation coefficients for each set of hyper-parameters were larger than $r = 0.99$. The correlations decreased in terms of 95% CrI widths. Specifically, 95% CrI widths produced by the informative prior were strongly
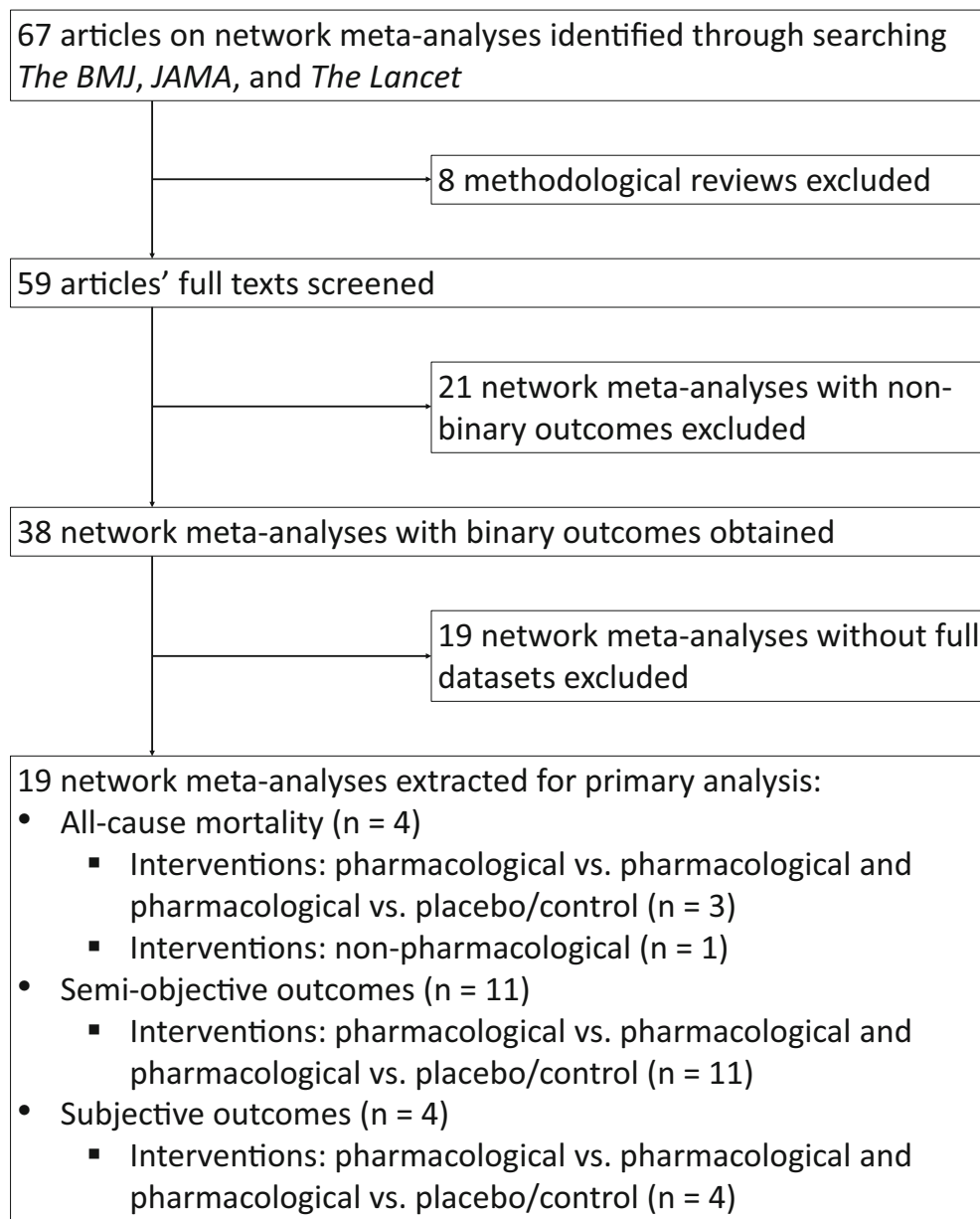
67 articles on network meta-analyses identified through searching *The BMJ*, *JAMA*, and *The Lancet*

8 methodological reviews excluded

59 articles' full texts screened

21 network meta-analyses with non-binary outcomes excluded

38 network meta-analyses with binary outcomes obtained

19 network meta-analyses without full datasets excluded

19 network meta-analyses extracted for primary analysis:
- All-cause mortality (n = 4)
  - Interventions: pharmacological vs. pharmacological and pharmacological vs. placebo/control (n = 3)
  - Interventions: non-pharmacological (n = 1)
- Semi-objective outcomes (n = 11)
  - Interventions: pharmacological vs. pharmacological and pharmacological vs. placebo/control (n = 11)
- Subjective outcomes (n = 4)
  - Interventions: pharmacological vs. pharmacological and pharmacological vs. placebo/control (n = 4)

Fig. 1 Flow diagram of network meta-analysis selection.

correlated with those by inverse-gamma priors $IG(0.1, 0.1)$, $IG(0.01, 0.01)$, and $IG(0.001, 0.001)$, all having $r = 0.90$. The half-normal priors, with $r = 0.89$ for $HN(0, 0.5)$, $r = 0.91$ for $HN(0, 1)$, and $r = 0.94$ for $HN(0, 2)$, also displayed strong correlations. The correlation with those by the uniform priors experienced greater variability, with $r = 0.87$ for $U(0, 2)$, $r = 0.82$ for $U(0, 5)$, and $r = 0.80$ for $U(0, 10)$. All P values of the above correlations were < 0.001.

Figure S5 in the Supplementary Material (Appendix C) presents the Bland–Altman plots among all 18 NMAs. It indicates strong agreement for both posterior median log ORs and 95% CrI widths by the various priors. Large differences in posterior median log ORs were likely observed when

the log ORs were close to 0, and large differences in 95% CrI widths were likely observed when CrIs were very wide.

Table 3 shows the kappa statistics between significant treatment comparisons identified via informative priors and non-informative priors. A total of 942 treatment comparisons were assessed among all NMAs. The kappa statistic for each pair of informative and non-informative priors was positive, mostly close to 1. A total of 236 treatment comparisons were found to be statistically significant via informative priors, which were more than those via non-informative priors.

Compared with the results based on the informative priors, the greatest variability in kappa statistics due to differences in hyper-parameters was observed when using the inverse-

Table 2 Summaries of the 19 Network Meta-analyses

Network meta-analysis	No. of studies	No. of treatments	Sample size	Outcome	Method	Original prior in primary analyses[a]	Informative prior[b]
Alfirevic 2015	95	13	20,293	Failure to achieve vaginal delivery within 24 h	Bayesian	$U(0,5)$	$LN(-3.23, 1.88^2)$; $LN(-3.02, 1.85^2)$
Anothaisintawee 2011	9	8	792	Decrease in National Institutes of Health chronic prostatitis symptom index	Frequentist	NA	$LN(-2.34, 1.62^2)$; $LN(-2.13, 1.58^2)$
Castellucci 2013	12	11	11,899	Recurrent venous thromboembolism	Bayesian and frequentist	$U(0,2)$	$LN(-3.23, 1.88^2)$; $LN(-3.02, 1.85^2)$
Castellucci 2014	22	3	10,600	Recurrent venous thromboembolism	Bayesian and frequentist	$U(0,5)^c$	$LN(-3.23, 1.88^2)$
Chatterjee 2013	21	8	23,113	Beta blockers effect on mortality	Bayesian and frequentist	$U(0,5)^c$	$LN(-4.27, 1.48^2)$; $LN(-4.06, 1.45^2)$
Cipriani 2018	473	22	111,282	Treatment of unipolar major depressive disorder	Bayesian	$U(0,5)$	$LN(-3.23, 1.88^2)$; $LN(-3.02, 1.85^2)$
Daniels 2012	18	7	2831	Rates of amenorrhea	Frequentist	NA	$LN(-3.23, 1.88^2)$
Dulai 2016	14	14	10,820	Any neoplasia	Bayesian	$U(0,5)$	$LN(-3.23, 1.88^2)$; $LN(-3.02, 1.85^2)$
Giacoppo 2015	13	3	2417	Target lesion revascularization	Bayesian	$U(0,2)$	$LN(-3.23, 1.88^2)$
Hazlewood 2016	28	23	10,537	American College of Rheumatology 50% improvement response: methotrexate-naïve	Bayesian	$U(0, 2)$	$LN(-3.23, 1.88^2)$
Isayama 2016	20	6	4455	Bronchopulmonary dysplasia	Bayesian	NA[d]	$LN(-3.23, 1.88^2)$
Palmerini 2015	10	7	31,666	All-cause mortality	Bayesian and frequentist	NA[d]	$LN(-3.93, 1.51^2)$
Phung 2010	13	6	6154	Achieved hemoglobin A_{1c} goal $< 7\%$	Bayesian and frequentist	NA[d]	$LN(-2.34, 1.62^2)$; $LN(-2.13, 1.58^2)$
Price 2014	28	4	6861	All-cause mortality	Bayesian	NA[d]	$LN(-4.27, 1.48^2)$; $LN(-4.06, 1.45^2)$
Siontis 2018	18	6	11,246	Invasive coronary angiography	Frequentist	NA	$LN(-3.23, 1.88^2)$; $LN(-3.02, 1.85^2)$
Stegeman 2018	9	11	11,111	Venous Thrombosis	Frequentist	NA	$LN(-3.23, 1.88^2)$; $LN(-3.02, 1.85^2)$
Wu 2013	61	11	36,772	All-cause mortality	Bayesian	$U(0,2)$	$LN(-4.27, 1.48^2)$; $LN(-4.06, 1.45^2)$
Xu 2018	23	7	12,621	Safety of immune checkpoint inhibitor drugs for cancer	Bayesian	NA[d]	$LN(-3.23, 1.88^2)$; $LN(-3.02, 1.85^2)$
Zheng 2018	7	3	69,959	Pancreatitis	Bayesian and frequentist	$U(0,5)^c$	$LN(-2.34, 1.62^2)$; $LN(-2.13, 1.58^2)$

[a]*Prior distribution for heterogeneity is only available for Bayesian analyses; it is not available (NA) for frequentist analyses*
[b]*Priors assigned based on the criteria by Turner et al.[14] In cells with two priors listed, the second was used for the secondary analyses because a placebo or control treatment is included in the network*
[c]*Prior was not specified in the published article, but was obtained by contacting the corresponding author*
[d]*The corresponding author was contacted for heterogeneity prior but did not respond at the time of submission*

gamma prior. Specifically, based on $IG(0.1, 0.1)$, there were 194 significant treatment comparisons identified with $\kappa = 0.87$. This number increased to 218 ($\kappa = 0.95$) using $IG(0.01, 0.01)$ and further increased to 232 ($\kappa = 0.97$) using $IG(0.001, 0.001)$. While $IG(0.001, 0.001)$ produced the strongest agreement with the informative priors, it also produced 5 treatment comparisons that were identified as significant by the non-informative priors but non-significant by the informative prior. All uniform priors had $\kappa = 0.93$ as did the half-normal priors $HN(0, 1)$ and $HN(0, 2)$; both uniform priors $U(0, 2)$ and $U(0, 5)$ produced 211 significant treatment comparisons, while this number slightly increased to 212 when using $U(0, 10)$. The half-normal priors $HN(0, 1)$ and $HN(0, 2)$

both produced 213 significant treatment comparisons; $HN(0, 0.5)$ yielded $\kappa = 0.94$ with 216 significant ones.

Impact of Priors Within Network Meta-analyses

Table 4 presents the correlation coefficients between the results and the kappa statistics within NMAs. The largest NMA (in terms of sample size) was Cipriani 2018, which included 111,282 samples, while the smallest NMA was Anothaisintawee 2011 with 792 samples. In the NMA of Cipriani 2018, the correlations of posterior median log ORs between the informative prior and each of the non-informative priors were nearly perfect with no discernable difference; they were all > 0.99 with P values < 0.001. An almost identical
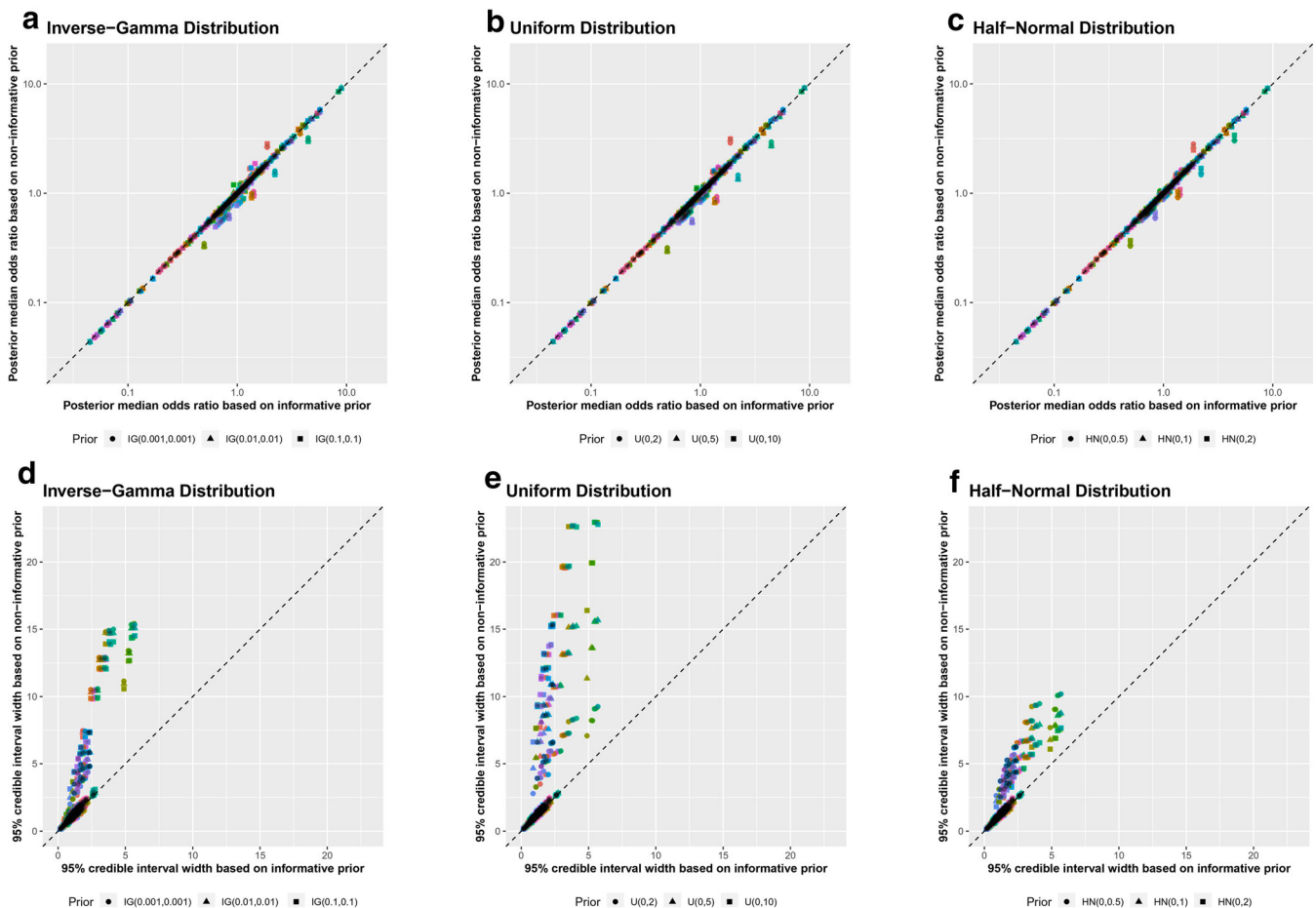
Fig. 2 Distributions of posterior median odds ratios on a logarithmic scale (a–c) and 95% credible interval widths (d–f) by various prior distributions among 18 network meta-analyses.

result was observed for the correlation between the 95% CrI widths.

The correlations between posterior median log ORs for each set of non-informative priors were consistently strong in the NMA of Anothaisintawee 2011; all non-informative priors for the median log ORs had correlation coefficients of at least 0.99. However, as this NMA had the smallest sample size, the correlations between the 95% CrI widths

exhibited more variability across different prior types and hyper-parameters; the half-normal priors had the highest correlation coefficients with values of 1.00, 0.99, and 0.97 for $HN(0, 0.5)$, $HN(0, 1)$, and $HN(0, 2)$, respectively. The least variability was observed for the inverse-gamma distribution; the correlation coefficients were 0.92 for $IG(0.1, 0.1)$ and 0.91 for both $IG(0.01, 0.01)$ and $IG(0.001, 0.001)$. The greatest variability across hyper-parameters was observed for the uniform prior; the priors $U(0, 2)$, $U(0, 5)$, and $U(0, 10)$ had correlation coefficients of 0.98, 0.91, and 0.89, respectively. All correlations had P values < 0.001.

A total of 231 treatment comparisons were produced to assess the agreement among priors in the NMA of Cipriani 2018. All hyper-parameters led to 100% agreement likely due to the large number of samples. On the other hand, the small NMA of Anothaisintawee 2011 contained a total of 28 treatment comparisons. The informative prior led to 2 significant comparisons, while all non-informative priors produced no significant comparison. This resulted in a kappa statistic that was incalculable.

Regardless of the size of NMAs, the informative and non-informative priors produced fairly similar point estimates of ORs, because all correlation coefficients were > 0.90. However,

Table 3 Kappa Statistics for Assessing the Agreement Between Informative and Non-Informative Priors with Respect to Significant Treatment Comparisons

Prior	Kappa	No. of significant treatment comparisons (percentage)
Informative prior	NA	236 (25.05%)
$IG(0.1, 0.1)$	0.87	194 (20.59%)
$IG(0.01, 0.01)$	0.95	218 (23.14%)
$IG(0.001, 0.001)$	0.97	232 (24.63%)
$U(0, 2)$	0.93	211 (22.40%)
$U(0, 5)$	0.93	211 (22.40%)
$U(0, 10)$	0.93	212 (22.51%)
$HN(0, 0.5)$	0.94	216 (22.93%)
$HN(0, 1)$	0.93	213 (22.61%)
$HN(0, 2)$	0.93	213 (22.61%)

Table 4 Correlation Coefficients Between the Results (Posterior Median Log Odds Ratios and 95% Credible Interval Widths) and Kappa Statistics for Assessing the Agreement Produced By Informative Priors and Non-Informative Priors Within Each Network Meta-analysis

Network meta-analysis	Estimate	Inverse-gamma prior[a]			Uniform prior[b]			Half-normal prior[c]		
		IG1	IG2	IG3	U1	U2	U3	HN1	HN2	HN3
Alfirevic 2015	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.97	0.97	0.97	0.95	0.95	0.95	0.95	0.95	0.97
Anothaisintawee 2011	Correlation: median log OR	0.99	0.99	1.00	0.99	0.99	0.99	1.00	1.00	0.99
	Correlation: 95% CrI width	0.92	0.91	0.91	0.98	0.91	0.89	1.00	0.99	0.97
	Kappa	NA	NA	NA	NA	NA	NA	NA	NA	NA
Castellucci 2013	Correlation: median log OR	0.90	0.97	1.00	0.94	0.93	0.93	0.97	0.96	0.95
	Correlation: 95% CrI width	0.97	0.97	0.98	0.97	0.96	0.95	0.98	0.98	0.97
	Kappa	NA	NA	NA	NA	NA	NA	NA	NA	NA
Castellucci 2014	Correlation: median log OR	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Chatterjee 2013	Correlation: median log OR	0.91	0.93	0.94	0.93	0.93	0.93	0.93	0.93	0.93
	Correlation: 95% CrI width	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.20	0.53	0.79	0.53	0.53	0.53	0.53	0.53	0.53
Cipriani 2018	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Daniels 2012	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.81	1.00	1.00	0.81	0.81	0.90	0.90	0.90	0.90
Dulai 2016	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.76	0.91	0.89	0.94	0.94	0.94	0.94	0.94	0.94
Giacoppo 2015	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	0.93	0.96	0.95	0.94	0.94	0.92	0.97	0.93	0.88
	Kappa	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Hazlewood 2016	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.79	0.93	0.99	0.90	0.90	0.90	0.91	0.91	0.91
Isayama 2016	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.00	0.63	1.00	0.63	0.63	0.63	0.63	0.63	0.63
Palmerini 2015	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	0.87	0.96	0.99	0.90	0.89	0.89	0.96	0.94	0.92
	Kappa	NA	NA	NA	NA	NA	NA	NA	NA	NA
Phung 2010	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Price 2014	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.40	0.40	1.00	0.67	0.67	0.67	0.67	0.67	0.67
Siontis 2018	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.87	0.87	1.00	0.87	0.87	0.87	1.00	0.87	0.87
Stegeman 2018	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.85	1.00	1.00	0.89	0.89	0.89	0.92	0.89	0.89
Xu 2018	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	0.90	1.00	0.90	0.91	0.91	0.91	0.91	0.91	0.91
Zheng 2018	Correlation: median log OR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Correlation: 95% CrI width	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Kappa	NA	NA	NA	NA	NA	NA	NA	NA	NA

[a]*Inverse-gamma priors for τ^2 with three sets of hyper-parameters: IG1, IG(0.1, 0.1); IG2, IG(0.01, 0.01); IG3, IG(0.001, 0.001)*
[b]*Uniform priors for τ with three sets of hyper-parameters: U1, U(0, 2); U2, U(0, 5); U3, U(0, 10)*
[c]*Half-normal priors for τ with three sets of hyper-parameters: HN1, HN(0, 0.5); HN2, HN(0, 1); HN3, HN(0, 2)*

the correlation between the 95% CrI widths for the different priors was possibly smaller and exhibited greater variability than that between point estimates of ORs in some NMAs. The informative priors typically produced narrower 95% CrIs than the non-informative priors (Fig. 2d–f). The greatest variability in correlations between 95% CrI widths was observed in Palmerini 2015 and in the smallest NMA of Anothaisintawee 2011. While all non-informative priors in Castellucci 2013 ($0.95 \leq r \leq 0.98$), Giacoppo 2015 ($0.88 \leq r \leq 0.97$), and Palmerini 2015 ($0.89 \leq r \leq 0.99$) had strong correlations, the correlations were not as strong as those in most other NMAs, which had $r > 0.99$ for all non-informative priors. Chatterjee 2013, Daniels 2012, Dulai 2016, Hazlewood 2016, and Phung 2010 had $r > 0.99$ for all priors except $IG(0.1, 0.1)$.

The Supplementary Material includes the scatterplots of posterior median ORs and 95% CrI widths by all priors (Figs. S6–S23 in Appendix C) and Bland–Altman plots (Figs. S24–S41 in Appendix C) in all NMAs separately.

There was generally small variability in kappa statistics across informative priors within NMAs. The four NMAs of Castellucci 2014, Cipriani 2018, Giacoppo 2015, and Phung 2010 had $\kappa = 1$ for all priors. For the NMAs of Castellucci 2013, Palmerini 2015, and Zheng 2018, each treatment comparison was non-significant by both the informative and non-informative priors. Xu 2018 had the same kappa statistic for all uniform and half-normal priors ($\kappa = 0.91$) and for $IG(0.1, 0.1)$ and $IG(0.001, 0.001)$ ($\kappa = 0.90$), while $IG(0.01, 0.01)$ had $\kappa = 1$. For Isayama 2016, $IG(0.1, 0.1)$ had $\kappa = 0$, $IG(0.001, 0.001)$ had $\kappa = 1$, and all other priors had identical agreement ($\kappa = 0.63$). Chatterjee 2013 ($\kappa = 0.53$), Dulai 2016 ($\kappa = 0.94$), and Price 2014 ($\kappa = 0.67$) all had identical agreements for the uniform and half-normal priors with different hyper-parameters. The greatest variability in agreements between hyper-parameters occurred when using the inverse-gamma prior; it was consistent with the observations in the overall assessment among all NMAs.

Secondary Analyses

A total of 11 NMAs contained a placebo or control treatment; the secondary analyses were performed for them by additionally considering alternative informative priors (i.e., for the type of comparisons with placebo/control as in Tables 1 and 2). Tables S2 and S3 and Figs. S42–S65 in the Supplementary Material (Appendix D) present the results. As in the primary analyses, the correlation coefficients of posterior median log ORs for each set of hyper-parameters were larger than $r = 0.99$. The correlations of 95% CrI widths for the secondary analyses were higher than their primary counterparts for $U(0, 10)$ ($r = 0.80$) and $HN(0, 2)$ ($r = 0.94$); all other correlations were lower in the secondary analyses than in the primary analyses. As in the primary analyses, the half-normal priors led to the highest correlation, and the greatest variability was observed for the uniform priors. The greatest variability in kappa statistics due to differences in hyper-parameters was observed when using the inverse-gamma prior, and the largest number of significant comparisons was observed using $IG(0.001, 0.001)$. All uniform and half-normal priors led to $\kappa = 0.95$. Four NMAs had $r > 0.99$ for all hyper-parameters for both the 95% CrI widths and median log ORs. All NMAs had strong correlations for all hyper-parameters; the weakest correlation was present in Anothaisintawee 2011, as in the primary analyses.

DISCUSSION

Main Findings

In this empirical study of 19 NMAs, we found that posterior median ORs produced by different priors had a very strong association. Noticeable variability appeared in estimates by different priors for NMAs with relatively small sample sizes per treatment comparison; thus, these NMAs tended to be sensitive to the prior specification. For large NMAs, non-informative priors generally produced nearly identical point estimates and 95% CrI widths, thus leading to an almost perfect agreement with the results based on informative priors. For small NMAs, the point estimates by informative priors were approximately the same as those produced by non-informative priors, but the CrIs produced by non-informative priors were often substantially wider than those produced by the informative priors. As a result, overall, informative priors yielded more statistically significant treatment effects. The greatest variability in agreement was observed when using the inverse-gamma priors, while the uniform and half-normal priors yielded approximately similar results.

Strengths and Limitations

This study considered most commonly used prior choices for modeling heterogeneity in the current practice of Bayesian NMAs. The results were based on recent NMAs published in high-impact medical journals, which were thus expected to be of high quality. All R code is provided in the Supplementary Material (Appendix E).

Nevertheless, this study had several limitations. First, we focused on assessing the impact of priors on the posterior medians and 95% CrIs of ORs, while the conclusions may not be directly generalized to other effect measures. Second, because the datasets involved in this study were relatively large, we did not examine the validity of several important assumptions (e.g., transitivity, consistency, reporting bias) in each NMA.[30] These factors may also influence NMA estimates along with the choice of priors, and researchers should investigate them on a case-by-case basis. Third, the NMAs published in high-impact journals may contain more studies, treatments, and samples than those published in other journals. The results of much smaller NMAs were likely more sensitive to prior choices.

Implications

Contemporary Bayesian NMAs published in high-impact journals do not adequately report details on the choice of heterogeneity prior distributions and their rationales. If an NMA does not have a large sample size, sensitivity analyses are recommended to examine the impact of using different hyper-parameters or other types of prior distributions, especially for the inverse-gamma prior. When the number of studies included in NMA is large, various non-informative priors produce similar conclusions. When the number of studies is small, conclusions become more sensitive to the prior type and hyper-parameters. In such cases, empirical informative priors may be used to produce more precise estimates.

Corresponding Author: *Lifeng Lin, PhD; Department of Statistics, Florida State University, Tallahassee, FL 32306, USA (e-mail: linl@stat.fsu.edu).*

Compliance with Ethical Standards:

This article focused on statistical methods for network meta-analyses, and all analyses were performed based on published data. Therefore, this study did not require ethical approval and patient consent.

Conflict of Interest: *The authors declare that they do not have a conflict of interest.*

REFERENCES

1. **Salanti G.** Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* 2012;3(2):80-97.

2. **Cipriani A**, **Higgins JPT**, **Geddes JR**, **Salanti G**. Conceptual and technical challenges in network meta-analysis. *Annals of Internal Medicine* 2013;159(2):130-37.

3. **Riley RD**, **Jackson D**, **Salanti G**, **Burke DL**, **Price M**, **Kirkham J**, **White IR**. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* 2017;358:j3932.

4. **Lumley T.** Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine* 2002;21(16):2313-24.

5. **Lu G**, **Ades AE**. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004;23(20):3105-24.

6. **Jackson D**, **White IR**, **Price M**, **Copas J**, **Riley RD**. Borrowing of strength and study weights in multivariate and network meta-analysis. *Statistical Methods in Medical Research* 2017;26(6):2853-68.

7. **Lin L**, **Xing A**, **Kofler MJ**, **Murad MH**. Borrowing of strength from indirect evidence in 40 network meta-analyses. *Journal of Clinical Epidemiology* 2019;106:41-49.

8. **Higgins JPT**, **Thompson SG**, **Deeks JJ**, **Altman DG**. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-60.

9. **Mills EJ**, **Thorlund K**, **Ioannidis JPA**. Demystifying trial networks and network meta-analysis. *BMJ* 2013;346:f2914.

10. **Nikolakopoulou A**, **Chaimani A**, **Veroniki AA**, **Vasiliadis HS**, **Schmid CH**, **Salanti G**. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLOS ONE* 2014;9(1):e86754.

11. **Carlin BP**, **Louis TA**. Bayesian Methods for Data Analysis. 3rd ed. Boca Raton, FL: CRC Press 2009.

12. **Gelman A.** Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 2006;1(3):515-34.

13. **Lu G**, **Ades AE**. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009;10(4):792-805.

14. **Turner RM**, **Davey J**, **Clarke MJ**, **Thompson SG**, **Higgins JPT**. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* 2012;41(3):818-27.

15. **Hutton B**, **Salanti G**, **Caldwell DM**, **Chaimani A**, **Schmid CH**, **Cameron C**, **Ioannidis JPA**, **Straus S**, **Thorlund K**, **Jansen JP**, **Mulrow C**, **Catalá-López F**, **Gøtzsche PC**, **Dickersin K**, **Boutron I**, **Altman DG**, **Moher D**. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine* 2015;162(11):777-84.

16. **Faltinsen EG**, **Storebø OJ**, **Jakobsen JC**, **Boesen K**, **Lange T**, **Gluud C**. Network meta-analysis: the highest level of medical evidence? *BMJ Evidence-Based Medicine* 2018;23(2):56-59.

17. **Tonin FS**, **Borba HH**, **Leonart LP**, **Mendes AM**, **Steimbach LM**, **Pontarolo R**, **Fernandez-Llimos F**. Methodological quality assessment of network meta-analysis of drug interventions: implications from a systematic review. *International Journal of Epidemiology* 2019;48(2):620-32.

18. **Gelman A**, **Carlin JB**, **Stern HS**, **Dunson DB**, **Vehtari A**, **Rubin DB**. Bayesian Data Analysis. 3rd ed. Boca Raton, FL: CRC Press 2014.

19. **Jansen JP**, **Crawford B**, **Bergman G**, **Stam W**. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. *Value in Health* 2008;11(5):956-64.

20. **Greco T**, **Landoni G**, **Biondi-Zoccai G**, **D'Ascenzo F**, **Zangrillo A**. A Bayesian network meta-analysis for binary outcome: how to do it. *Statistical Methods in Medical Research* 2016;25(5):1757-73.

21. **Spiegelhalter DJ**, **Abrams KR**, **Myles JP**. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester, UK: John Wiley & Sons 2004.

22. **Friede T**, **Röver C**, **Wandel S**, **Neuenschwander B**. Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods* 2017;8(1):79-91.

23. **Zhang J**, **Carlin BP**, **Neaton JD**, **Soon GG**, **Nie L**, **Kane R**, **Virnig BA**, **Chu H**. Network meta-analysis of randomized clinical trials: reporting the proper summaries. *Clinical Trials* 2014;11(2):246-62.

24. **Dias S**, **Sutton AJ**, **Ades AE**, **Welton NJ**. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making* 2013;33(5):607-17.

25. **Gelman A**, **Rubin DB**. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992;7(4):457-72.

26. **Kass RE**, **Carlin BP**, **Gelman A**, **Neal RM**. Markov Chain Monte Carlo in practice: a roundtable discussion. *The American Statistician* 1998;52(2):93-100.

27. **Cipriani A**, **Furukawa TA**, **Salanti G**, **Chaimani A**, **Atkinson LZ**, **Ogawa Y**, **Leucht S**, **Ruhe HG**, **Turner EH**, **Higgins JPT**, **Egger M**, **Takeshima N**, **Hayasaka Y**, **Imai H**, **Shinohara K**, **Tajika A**, **Ioannidis JPA**, **Geddes JR**. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet* 2018;391(10128):1357-66.

28. **Cohen J.** A coefficient of agreement for nominal scales. *Educational and Psychosocial Measurement* 1960;20(1):37-46.

29. **Landis JR**, **Koch GG**. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.

30. **Puhan MA**, **Schünemann HJ**, **Murad MH**, **Li T**, **Brignardello-Petersen R**, **Singh JA**, **Kessels AG**, **Guyatt GH**. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349:g5630.