HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2020 May 29.

Published in final edited form as:

Stat Med. 2018 September 30; 37(22): 3254–3266. doi:10.1002/sim.7815.

Multivariate network meta-analysis to mitigate the effects of outcome reporting bias

Hyunsoo Hwang,

Department of Biostatistics and Data Science, The University of Texas School of Public Health, TX 77030, USA

Stacia M. DeSantis

Department of Biostatistics and Data Science, The University of Texas School of Public Health, TX 77030, USA

Abstract

Outcome reporting bias (ORB) is recognized as a threat to the validity of both pairwise and network meta-analysis (NMA). In recent years, multivariate meta-analysis (MMA) methods have been proposed to reduce the impact of ORB in the pairwise setting. These methods have shown that MMA can reduce bias and increases efficiency of pooled effect sizes. However, it is unknown whether multivariate NMA (MNMA) can similarly reduce ORB. Additionally, it is quite challenging to implement MNMA due to the fact that correlation between treatments and outcomes must be modeled, thus the dimension of the covariance matrix and number of components to estimate grows quickly with the number of treatments and number of outcomes. To determine whether MNMA can reduce the effects of ORB on pooled treatment effect sizes, we present an extensive simulation study of a Bayesian MNMA. Via simulation studies, we show that MNMA reduces the bias of pooled effect sizes under a variety of outcome missingness scenarios, including missing at random and missing not at random. Further, MNMA improves the precision of estimates, producing narrower credible intervals. We demonstrate the applicability of the approach via application of MNMA to a multi-treatment systematic review of randomized controlled trials of anti-depressants for the treatment of depression in older adults.

Keywords

meta-analysis; multivariate meta-analysis; network meta-analysis; outcome reporting bias; publication bias

1. Introduction

Network meta-analysis (NMA), often called mixed (or multiple) treatment comparisons (MTC) meta-analysis, has been increasingly studied in recent years.[1] Interest is due to the fact that systematic reviews, which are considered the pinnacle of evidence-based medicine, often result in “networks of evidence” in which many different treatments have been

Supplementary material

Additional simulation results, simulation procedure, and OpenBUGS code are available in the Supplementary Materials.

compared. [2,3] These evidence structures are composed of both direct evidence (e.g., B:C trials) and indirect evidence, which is obtained through a common comparator trials (e.g., evidence for B:C synthesized via trials comparing A:B and A:C). [2,4] NMA can be applied to these structures to generate pooled (weighted mean) estimates of effect sizes, even for treatments that have never been directly compared. To this end, several Bayesian and frequentist methods for NMA exist for analysis including arm-based parameterizations [5] and difference-or contrast-based parameterizations of pooled effect sizes. [2–4,6–8]

Very recently, researchers have turned attention toward the issue of missing data in networks of evidence. Reporting bias (RB), including outcome reporting bias (ORB) and publication bias (PB), are particular challenges that have been discussed in pairwise meta-analysis, but less so in network settings. [9–20] Outcome reporting bias occurs when an outcome is not published based on its significance or direction in a given study. A study is subject to publication bias (PB) when none of the outcomes in the study is reported based on some characteristic, usually the significance, direction, or size of the study (i.e., the entire study is unreported). In a univariate analysis, ORB is essentially equivalent to PB. The focus of the current paper is the setting of ORB.

To exemplify the extent of the problem that ORB presented in network meta-analysis, a study in Journal of the American Medical Association (JAMA) Psychiatry estimated that 33% of trials of second-generation anti-depressants registered with the Food and Drug Administration (FDA) are either not published, or do not publish all pre-specified outcomes. [12] A number of other publications have also drawn attention to the statistical bias that afflicts mental health research, though the issue is certainly not confined to this field. [21–24] The issue is of particular concern when only aggregate data (versus patient-level data) are available, the most common setting where meta-analysis is applied.

In the pairwise treatment setting, multivariate meta-analysis (MMA) has been shown to reduce bias and increase efficiency of pooled effect sizes for outcomes subject to ORB. [24–27] However, a principled way to apply multivariate NMA (MNMA) to mixed treatments comparisons in a network of trials has not been well-studied. A procedure must model the covariance between treatments within studies, and outcomes within and between studies. Liu et al [8] recently presented a multivariate NMA model that uses a Clayton copula distribution to model correlated binary outcomes and show via simulation studies that modeling correlated outcomes reduces bias in pooled log odds ratios (LORs). Two recent papers presented bivariate NMA approaches applied to the same network of acute mania, [28,29] while a third applied multivariate NMA to a systematic review of poison prevention strategies with 3 outcomes. [30] Specifically, Efthimiou et al [29] extensively investigated approaches to decompose the complex variance-covariance matrix required for modeling treatment and outcome correlation, essentially minimizing the number of parameters in the matrix to ease computational burden. Very recently, Jackson et al [31] proposed a MNMA model which uses a matrix-based method of moments estimators, which is advantageous for relatively fast computation time, compared with other models based on MCMC and REML. Unlike other approaches, Hong et al [5,32] employed an arm-based NMA model, in which the absolute rather than relative treatment effects are pooled, and extend this approach to accommodate multiple outcomes. In the current paper, we use the contrast-based approach

due to the fact that the arm-based approach could violate the randomization principle and bias the estimates of relative treatment effects with inflated posterior variance, as discussed by Dias and Ades.[33]

Given the computational and programming burden of fitting a multivariate NMA, it is important to determine their practical value for researchers; however, few studies have assessed the utility and robustness of MNMA to ORB, thus its practical value remains unstudied and leaves researchers with many meta-analytic tools from which to choose without guidance. A comprehensive simulation study supported by case studies are needed to determine whether MNMA is capable of mitigating effects of ORB in the network setting and are the focus of this paper.

We apply the method of Efthimiou et al [29] to consider "true" multivariate outcomes (i.e., > 2 outcomes) and perform an extensive simulation study testing the ability of MNMA to reduce ORB in a variety of missingness settings where outcomes (L) > 2. As a case study, we consider a mental health application – a systematic review of randomized controlled trials assessing the effect of anti-depressants for treatment of depression in older adults. [34]

The rest of the paper is outlined as follows. In Section 2, we present the multivariate NMA of Efthimiou et al [29] which encompasses three outcomes. In Section 3, we assess the robustness of MNMA to ORB. Using the nomenclature of Rubin,[35] we consider 3 outcome missingness scenarios; missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), under different outcome correlation structures. In Section 4, we apply the method to a systematic review of antidepressants used in the treatment of major depressive disorder for older adults. In Section 5, we conclude with a discussion.

2. Methods

We first briefly review a multivariate meta-analytic model for two treatments. We specifically present Riley's "reduced" MMA model where a single correlation coefficient is specified for within- and between- studies correlations. We then introduce the multivariate NMA model of Efthimiou et al,[29] which generalizes Riley's model for mixed treatments comparisons. Throughout, "multivariate" refers to bivariate or trivariate models while, "univariate" refers to models with a single outcome. Since the outcomes of interest are binary, the treatment effects are parameterized as log odds ratios (LORs), though the approaches are generalizable to continuous, time-to-event, or mixed outcomes. [29]

2.1. Multivariate meta-analysis model

Assuming all studies report all 3 outcomes with complete data (no missing outcomes reported), the MMA model is,

$$\begin{pmatrix} y_{1,1} \\ y_{1,2} \\ y_{1,3} \\ y_{2,1} \\ \vdots \\ y_{Ns,3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \gamma_{1,1} \\ \gamma_{1,2} \\ \gamma_{1,3} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{N_s,1} \end{pmatrix} + \begin{pmatrix} \in_{1,1} \\ \in_{1,2} \\ \in_{1,3} \\ \in_{2,1} \\ \vdots \\ \in_{N_s,1} \end{pmatrix}, \quad (1)$$

where $y_{i,j}$ is an observed relative effect (e.g., the log odds ratio) of the ith observation from total N_s number studies on outcome j, and the parameters β_1,β_2, and β_3 represent pooled (or mean) relative effects for each outcome. Equation (1) can be rewritten in matrix notation as, $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y}$ is a $3N_s$ dimensional vector and $\boldsymbol{X}$ is a design matrix of dimension $3N_s \times 3$. The model includes two random vectors, $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$, which account for the variation between studies and within a study, respectively. Specifically, the parameter $\gamma_{i,j}$ is a random effect for the variation between studies for outcome j, and the $\boldsymbol{\gamma}$ follows a multivariate normal distribution, $\boldsymbol{\gamma} \sim N(0, \boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ is a *between-study* variance- covariance matrix. The matrix $\boldsymbol{\Phi}$ is block diagonal, i.e. $\boldsymbol{\Phi} = \text{diag}(\Phi_1,\Phi_2,\ldots,\Phi_{Ns})$, and each block is written as,

$$\boldsymbol{\Phi}_i = \begin{pmatrix} \tau_1^2 & \rho_{12}^B \tau_1 \tau_2 & \rho_{13}^B \tau_1 \tau_3 \\ \rho_{12}^B \tau_1 \tau_2 & \tau_2^2 & \rho_{23}^B \tau_2 \tau_3 \\ \rho_{13}^B \tau_1 \tau_3 & \rho_{23}^B \tau_2 \tau_3 & \tau_3^2 \end{pmatrix}, \quad (2)$$

where. $\tau_1^2, \tau_2^2,$ and τ_3^2 are variances of heterogeneity for each outcome, and $\rho_{12}^B,\ \rho_{13}^B,$ *and* ρ_{23}^B are correlation coefficients between outcome 1 and 2, outcome 1 and 3, and outcome 2 and 3, respectively. The matrix, Φ_i, is assumed to be the same for each study i.

The random error, $\epsilon_{i,j}$, denotes the variation within a study; it is the sampling error in the ith study for the jth outcome. The vector, $\boldsymbol{\varepsilon}$, from Equation (1) is also assumed to follow multivariate normal distribution, $\boldsymbol{\varepsilon} \sim N(0,\boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is a *within-study* block diagonal variance covariance matrix and each component of $\boldsymbol{\Omega}$ for study i is written as,

$$\Omega_i = \begin{pmatrix} s_{i1}^2 & \rho_{i,12}^w s_{i1} s_{i2} & \rho_{i,13}^w s_{i1} s_{i3} \\ \rho_{i,12}^w s_{i1} s_{i2} & s_{i2}^2 & \rho_{i,23}^w s_{i2} s_{i3} \\ \rho_{i,13}^w s_{i1} s_{i3} & \rho_{i,23}^w s_{i2} s_{i3} & s_{i3}^2 \end{pmatrix}, \quad (3)$$

where s_{ij}^2 is the sampling error for study i and outcome j (the variance of the effect size y_{ij}). The vector $\left(\rho_{i,12}^w,\ \rho_{i,13}^w,\ \rho_{i,23}^w\right)$ are correlation coefficients within each study, i, between outcomes 1 and 2, outcomes 1 and 3, and outcomes 2 and 3, respectively. The challenge of this model is the lack of knowledge on within-study correlation coefficients, ρ_i^w, which needs to be specified. In general, few studies provide sufficient information for estimating these coefficients. [26] In an effort to resolve this problem, Riley et al [36] proposed a reduced model for multivariate outcomes that contains a single overall correlation coefficient to

replace both coefficients, ρ^B and ρ_i^w that does not require specification of the within-study correlation coefficients. In this case, $\boldsymbol{Y}$ can be modeled as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\nu}$, with $\boldsymbol{\nu} \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\nu}$ is assumed to follow a multivariate normal distribution and $\boldsymbol{\Sigma}$ is again a block-diagonal matrix. The overall variance-covariance matrix for study i is written as,

$$\Sigma_i = \begin{pmatrix} \psi_1^2 + s_{i1}^2 & \cdot & \cdot \\ \rho_{i,12}^g \sqrt{(\psi_1^2 + s_{i1}^2)(\psi_2^2 + s_{i2}^2)} & \psi_2^2 + s_{i2}^2 & \cdot \\ \rho_{i,13}^g \sqrt{(\psi_1^2 + s_{i1}^2)(\psi_3^2 + s_{i3}^2)} & \rho_{i,23}^g \sqrt{(\psi_2^2 + s_{i2}^2)(\psi_3^2 + s_{i3}^2)} & \psi_3^2 + s_{i3}^2 \end{pmatrix}, \quad (4)$$

where ρ^g is a global correlation coefficient, which is an amalgam of both correlation coefficients, ρ^B and ρ^w, and can be modeled identically across the studies ($\rho_i^g = \rho^g$). [29,36] We note that ψ is not identical to τ in equation (2) due to the inclusion of the global correlation coefficient ρ^g in Equation (4).

2.2. Multivariate network meta-analysis model

Efthimiou et al [29] extended Riley's MMA model to accommodate multiple treatment arms by decomposing the variance-covariance matrix to reflect the correlations between treatments and outcomes. In the multiple treatments setting, we consider N_T total treatments in a network where each study contains a maximum of 3 outcomes. The consistency equations of Lu and Ades [2] are assumed to hold for every outcome, implying that the vector of pooled effect sizes can be written as a function of basic parameters which are treatment effects relative to the reference treatment.

$$\beta_{(B:C),l} = \beta_{(A:C)l} - \beta_{(A:B)l} \text{ for} l = 1, 2, 3, \quad (5)$$

where the functional parameter, $\beta_{i,(B:C)l}$ is a pooled log odds ratio of treatment C relative to treatment B for outcome l in the ith study. The right hand side of Equation (5) represents basic parameters, assuming A is a reference treatment.

For 2-arm studies reporting 3 outcomes, the MNMA model is $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\nu}$,

$$\begin{pmatrix} y_{1,(A:B),1} \\ y_{1,(A:B),2} \\ y_{1,(A:B),3} \\ y_{2,(B:C),1} \\ y_{2,(B:C),2} \\ y_{2,(B:C),3} \\ y_{3,(A:C),1} \\ \vdots \\ y_{N_S,(A:C),3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \beta_{A:B,1} \\ \beta_{A:B,2} \\ \beta_{A:B,3} \\ \beta_{A:C,1} \\ \beta_{A:C,2} \\ \beta_{A:C,3} \end{pmatrix} + \begin{pmatrix} \nu_{1,(A:B),1} \\ \nu_{1,(A:B),2} \\ \upsilon_{1,(A:B),3} \\ \upsilon_{2,(B:C),1} \\ \upsilon_{2,(B:C),2} \\ \upsilon_{2,(B:C),3} \\ \upsilon_{3,(A:C),1} \\ \vdots \\ \upsilon_{N_S,(A:C),3} \end{pmatrix}, \quad (6)$$

where $\boldsymbol{Y}$ is the vector of observed effects (log odds ratios), $\boldsymbol{X}$ is a design matrix that represents all treatment contrasts in the network, [37] $\boldsymbol{\beta}$ is the $(N_T - 1) \times 3$ dimensional vector

of basic parameters, $\boldsymbol{\nu}$ is the combined vector of random errors and additional variations due to heterogeneity, with $\boldsymbol{\nu} \sim N(0, \boldsymbol{\Sigma})$. In this model, the heterogeneity is assumed to be constant between different comparisons. The main difference between MMA and MNMA is the design matrix $\boldsymbol{X}$ which maps the observed treatment comparisons, using basic parameters, and variance-covariance matrix, Σ_i. When there are only 2-arm studies in the network, the same variance-covariance matrix Σ_i, from Equation (4), can be employed. However, there are often studies containing more than 2 arms in a network. For the 3-arm study, the variance-covariance matrix expands to 6 × 6 matrix, which results in additional correlation coefficients between outcomes and treatments. To reduce the burden of additional parameter estimation, Efthimiou et al [29] substantially simplified the variance-covariance matrix. Based on the homogenous variance assumption, these authors transformed the matrix in a way that minimizes the number of parameters to estimate, which is easy to apply and reduce the computational burden. Details of the derivation for the simplified matrix is found in the Supplementary Materials of that paper.[29] The simplified variance-covariance matrix for 3-arms and 3 outcomes that compares treatments A, B, and C is given as,

$$\Sigma_i = \begin{pmatrix} \upsilon_{i,AB,1} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{12}^{g}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AB,2}} & \upsilon_{i,AB,2} & \cdot & \cdot & \cdot & \cdot \\ \rho_{13}^{g}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AB,3}} & \rho_{23}^{g}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AB,3}} & \upsilon_{i,AB,3} & \cdot & \cdot & \cdot \\ \frac{1}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,1}} & \frac{\rho_{12}^{g}}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,1}} & \frac{\rho_{13}^{g}}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,1}} & \upsilon_{i,AC,1} & \cdot & \cdot \\ \frac{\rho_{12}^{g}}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,2}} & \frac{1}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,2}} & \frac{\rho_{23}^{g}}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,2}} & \rho_{12}^{g}\sqrt{\upsilon_{i,AC,1}\upsilon_{i,AC,2}} & \upsilon_{i,AC,2} & \cdot \\ \frac{\rho_{13}^{g}}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,3}} & \frac{\rho_{23}^{g}}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,3}} & \frac{1}{2}\sqrt{\upsilon_{i,AB,3}\upsilon_{i,AC,3}} & \rho_{13}^{g}\sqrt{\upsilon_{i,AC,1}\upsilon_{i,AC,3}} & \rho_{23}^{g}\sqrt{\upsilon_{i,AC,2}\upsilon_{i,AC,3}} & \upsilon_{i,AC,3} \end{pmatrix}, \quad (7)$$

where $\upsilon_{i,AB,1}$ represents $\psi_1^2 + s_{i1}^2$ for outcome 1 in the ith study between treatment A and B comparison. With more than 2 outcomes, there is a singular estimated matrix problem. To ensure the variance-covariance matrices, (4) and (7), are positive definite, the Cholesky decomposition is used. [38]

In Efthimiou et al [29], only bivariate outcomes were considered although the approach for 3 outcomes was outlined in Supplementary Materials. The depression networks in our study contain 2 and 3 arm trials reporting on 3 outcomes; accordingly, we apply the model of Efthimiou et al [29] to accommodate 3 outcomes. Our extensive simulation study also assesses the robustness of the model to ORB for 3 outcomes.

2.3. Bayesian modelling

The MNMA model presented in equations (5–7) is fit within a Bayesian framework using Markov chain Monte Carlo (MCMC) methods, which can flexibly combine evidence from multiple sources by incorporating prior beliefs, and reduce complexity when there are missing outcomes. When summarizing evidence from multiple outcomes, the posterior distribution of missing values can be estimated directly from the model through MCMC with predictive distributions. [39] For simulation and data analysis, we initialize 3 chains and assess convergence using trace plots, density plots, and Rubin & Gelman convergence diagnostic. [40] For prior distributions of effect sizes and correlation coefficients, we assign vague (or non-informative) uniform priors, ~N(0,10000) and $\rho^g \sim U(-1,1)$. We also assume that each correlation coefficient is common across studies ($\rho_i^g = \rho^g$). For the variance parameters, we choose weakly informative uniform priors, $\tau \sim U(0,2)$ and $\psi \sim U(0,2)$ for both univariate and multivariate NMA. We implement analysis in OpenBUGS (Ver.3.2.3.), and R (Ver 5.4.1.) and report posterior means, and 95% credible intervals. Statistical "significance" is achieved when a 95% credible interval excludes the null value.

3. Simulation Study

To evaluate the ability of the MNMA to reduce the impact of outcome reporting bias on pooled estimates, we compare the performance of MNMA to UNMA under 3 different missingness mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). [35] We adapt the approach from the previous studies. [8,24,32,41] We assume there are 5 treatment interventions (K=5) and 3 total outcomes (L=3) where at least one outcome is reported in all studies in the network. This assumption is reasonable for the motivating examples owing to the fact that at least one major outcome such as efficacy or discontinuation must be reported in an RCT publication. Each generated network consists of 50 studies, 84% of which are 2-arm and 16% of which are 3-arm studies, reflecting both the current application and networks of RCTs for prevalent mental health disorders, in which most studies are 2 arm but some 3 arm studies exist. [42,43] The numbers of treatments included in the simulation are approximately 29%, 21%, 17%, 17%, and 16% respectively for treatments A, B, C, D, and E. We set the probabilities of the event for treatment A, B, C, D, and E in outcome 1, 2, and 3 as follows. $(p_{A1},p_{B1},p_{C1},p_{D1},p_{E1})$=(0.35,0.4,0.5,0.55,0.6), $(p_{A2},p_{B2},p_{C2},p_{D2},p_{E2})$=(0.62,0.48,0.42,0.37,0.35), and $(p_{A3},p_{B3},p_{C3},p_{D3},p_{E3})$=(0.4,0.5,0.6,0.65,0.65) where p_{kl} denotes the probability of event for treatment k and outcome l. When the treatment has a protective effect compared to the reference treatment, A, the LOR for efficacy (the primary outcome) should be greater than 0, thus these choices result in clinically sensible effect sizes, e.g., $(lor_{AB1},lor_{AC1},lor_{AD1},lor_{AE1})$=(0.214,0.619,0.820,1.025),

(lor_{AB2},lor_{AC2},lor_{AD2},lor_{AE2})=(−0.570,−0.812,−1.021,−1.109), and (lor_{AB3},lor_{AC3},lor_{AD3},lor_{AE3})=(0.405,0.811,1.025,1.025). We set both heterogeneity standard deviations, τ and ψ, to 0.3. We assume the presence of binary outcome 2 is a harmful outcome that could be negatively correlated with outcome 1, which is why the LORs are negative for outcome 2. Consistent with the design of most frequentist RCTs, within a study, the sample size is assumed to be equal for each treatment arm, and the sample size for each study is drawn from a uniform distribution U(75, 125). Due to the limited space, we only illustrate comparisons between the reference treatment, A, and other treatments.

We conduct simulations under 3 outcome correlation structures - uncorrelated, moderate, and strong. The correlation of the binary data between outcome 1 and 2, outcome 1 and 3, and outcome 2 and 3 are one of, (0,0,0) (−0.5,0.5,−0.5), or (−0.8, 0.8, −0.8), respectively, which we generate using a copula model for binary outcomes from the R package, *copula.* [44] The Copula model allows for specification of the univariate margins and the multivariate dependence structure independently,[8] hence, margins can be correlated binomial distributions regardless of a different multivariate distribution such as a multivariate normal distribution. We assume outcome 1 is always reported (e.g., an outcome such as efficacy), but outcomes 2 and/or 3 may be missing. These outcomes may refer to measurements like discontinuation, dropout, medication adherence, or adverse events. We set the percentage of missingness to 30% or 50%. Under MCAR, the probability of missingness is simulated from a Bernoulli distribution. Under MAR, the probability of missingness is associated only with the observed outcome 1. Under MNAR, the probability of missingness is associated with unobserved outcomes, either outcome 2 or 3, or both. For MAR and MNAR, under which ORB would arise, we generate missingess according to the significance of outcome in a given study. We conduct a Fisher's exact test to determine the significance of the missingness. Under MNAR, outcome 2 should be less likely to be missing if outcome 2 in a given study is highly significant. The logic is similar for outcome 3. Under MAR, outcomes 2 and 3 of a given study would be less likely to be missing if the outcome 1 is highly significant. The details of the data generation procedure for MAR and MNAR outcomes are outlined in the Supplementary Materials.

As described above, we use non-informative priors for correlation coefficients ρ^g~U(−1,1), and for pooled relative effects, β~N(0,10000). We use a weakly informative prior for random effects, ψ~ (0,2). Under each simulation scenario, we generate 200 networks and apply UNMA and MNMA methods to each. We summarize the simulations in terms of bias in the posterior mean LOR and 95% coverage probability (CP) as determined by whether the 95% credible interval includes the true value. Owing to computation time for simulation in OpenBUGS, we obtain 5,000 samples after discarding first 5,000 samples as burn-in. The steps for the analysis can be summarized as:

i. Apply UNMA for each outcome separately using the prior distributions, τ~U(0,2) to each of the 4 missingness scenarios including complete data, MCAR, MAR, and MNAR under the 3 specified correlation structures (0, 0.5, 0.8).

ii. Apply MNMA to 3 outcomes using the following prior distributions, $\rho_{12}^{g} \sim U(-1,1)$, $\rho_{13}^{g} \sim U(-1,1)$, $\rho_{23}^{g} \sim U(-1,1)$, $\psi_1 \sim U(0,2)$, $\psi_2 \sim U(0,2)$, and $\psi_3 \sim U(0,2)$ under the complete data, MCAR, MAR, and MNAR under the 3 specified correlation structures (0, 0.5, 0.8).

Figures S1 and S2 in the Supplementary Materials summarize the results of the empirical bias and the CP of log odds ratios of UMNA and MMNA when the correlation between outcomes is zero. The solid blue line represents the bias in the LOR resulting from MNMA, and the dotted red line represents bias resulting from UNMA. As expected, when the correlation between outcomes is zero, there is no advantage of MNMA to UNMA in terms of bias reduction and the posterior estimates of parameters and coverage probabilities are nearly exactly the same. This comparison serves as a check of our MNMA formulation and coding.

Figures 1 and 2 compare the methods when the correlation among outcomes is moderate, i.e., $|\rho| = 0.5$, and the number of missing studies is 50%. First, as expected, there is little difference in performance between the two methods under the complete data or MCAR. However, under MAR missingness, the UNMA results in severely biased LORs for outcomes 2 and 3. Application of MNMA substantially reduces this bias (shown in columns 2 and 3 of Figure 1), as compared to UNMA, e.g., for $LOR132$ and $LOR133$. The CPs similarly demonstrate a difference in coverage between two methods. Figure 2 shows the coverage probabilities for the UNMA are as low as 0.89 and 0.90 for $LOR142$ and $LOR143$, as compared to 0.92 and 0.93, respectively from fitting the MNMA.

Figure 1 also shows that under MNAR, the bias is much larger compared to that under MAR, as would be expected. However, even under MNAR, MNMA results in smaller bias, as compared to UNMA. For example, the bias in $LOR132$ decreases from 0.20 to 0.13. The trends in CP are similar.

Contrary to bias reduction, bias increase is also observed by employing MNMA. In general, the LORs for outcome 1 (which is always observed), are estimated correctly by both approaches. However, minimally larger bias results from MNMA under the MNAR setting, compared with the UNMA. For example, the bias for $LOR141$ increases from 0.006 to 0.0135 using the multivariate approach, which is shown in the column 1 of the Figure 1.

Figures S3 and S4 compare the two methods under 30% missingness. As the proportion of missingness decreases, it is clear that bias in LORs is smaller and the corresponding CPs larger for missing outcomes regardless of the approach used; however, the MNMA still reduces bias and improves coverage probability substantially. For example, the bias in $LOR142$ decreases from 0.136 to 0.084 when using the multivariate approach for MNAR data. Figures S5–S8 in the Supplementary Materials compare the two methods when the correlation among outcomes is very strong, i.e., $|\rho| = 0.8$. As expected, under very strong outcome correlation, MNMA substantially reduces the bias in pooled effect sizes for outcomes 2 and 3 over UNMA (shown in columns 2 and 3 of the Figures S5 and S7), showing an obvious dose response relationship between reduction of bias as correlation

increases. The trends for CPs are also similar; these better achieve the nominal coverage rate as correlation increases. (See Figures S6 and S8)

4. Application to the antidepressant networks

Thorlund et al [34] present a systematic review and network meta-analysis of efficacy and safety of selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs) and other active antidepressants in older adults for the treatment of major depressive disorder. There is a total of 4,588 participants across 15 RCTs. We choose this study to illustrate the multivariate method due to amount of missingness present in the efficacy and safety data, and due to a reasonably high correlation between these outcomes. In the current study, three outcomes are included for multivariate analysis: response, all-cause discontinuation (also called dropout), and discontinuation due to adverse events. Response is defined as at least 50% reduction in Hamilton Depression Rating Scale score or Montgomery-Asberg Depression Rating Scale score from baseline. All-cause discontinuation and discontinuation due to AEs are defined as leaving the study for any reason, and leaving the study due to adverse events, respectively. Figure 3 present the network structure for each outcome. The size of the dots is proportional to the number of studies in the network including that treatment. A line exists if two treatments have been compared head-to-head while the thickness of the line is proportional to the number of times those treatments have been compared. The first panel shows that the network for treatment response is sparser than that for other outcomes; in fact, 12 (80%) studies report response, 15 (100%) report all-cause discontinuation, and 14 (93%) report discontinuation due to AEs. The number of studies per each intervention is summarized in Table S1. Note that diazepam, which is not an anti-depressant, was used as a placebo arm in one trial. In this application, settings for the prior distributions are identical to those in the simulation. For both univariate and multivariate methods, we obtain 150,000 samples after discarding first 10,000 samples as burn-in.

Table 1 presents the posterior median and mean estimates and 95% credible intervals for heterogeneity variances, τ^2/ ψ^2, and correlation parameters, ρ^g, respectively for the UNMA and MNMA methods. The posterior mean values for the correlation parameters in the variance-covariance matrix from Equation (4) demonstrate a moderate negative relationship between response and the other two outcomes, and a moderate positive relationship between all-cause discontinuation and discontinuation due to AEs. The median estimates of heterogeneity variances are slightly different between the two methods, likely because the network is small. Figure 4 shows the posterior mean ORs and 95% credible intervals for each outcome for placebo versus all other treatments. The presence of correlation and the fact that all-cause discontinuation is a completely observed outcome, leads to an improvement in precision of the 95% credible intervals for outcomes subject to missingness; the width of the 95% CIs resulting from the MNMA (solid line) are narrower compared to those resulting from the UNMA (dotted line), likely a result of borrowing information from the two additional outcomes. For each outcome, there is also a slight difference in the point estimates of the ORs between multivariate (dotted line) and univariate (solid line) approaches. The different point estimates and narrowed 95% CIs in some comparisons lead to “significant” results not otherwise observed via univariate analysis; e.g., the odds of

discontinuation due to AEs is significantly elevated in the venlafaxine vs placebo and fluoxetine vs placebo group (third panel in Figure 4). The odds of response in the duloxetine vs placebo is also significantly elevated. (first panel in Figure 4)

Figure 5 compares the multivariate (blue) versus univariate (red) posterior probabilities of each possible position ranked from 1 (best) to 10 (worst) for response rate. These figures are also available for all-cause discontinuation and discontinuation due to AEs; see Figures S9 and S10 in the Supplementary Material. Results from the multivariate approach indicate the ranking distribution of citalopram changes favorably versus the univariate approach, which is also consistent with the change in point estimates of the LOR (first panel, Figure 4). On the other hand, the ranking distributions for placebo and paroxetine indicate slightly lower response rankings are favored by the multivariate versus univariate approach; specifically, the ranking distribution for placebo indicates it is less effective in eliciting a response, as one would expect from a placebo. Note, the ranking distribution for trazodone cannot be interpreted. The one trial of trazadone in this network did not report a response. Since there is no available data for this treatment, samples from the MCMC algorithm represent normally distributed noise, leading to more extreme values in both directions for the ranking distribution. As expected, there is almost no change in the ranking distributions for the fully observed outcome, all-cause discontinuation (Figure S9 in the Supplementary Materials). There are slight non-notable changes in the ranking distributions for discontinuation due to AEs (Figure S10 in the Supplementary Materials).

5. Discussion

Outcome reporting bias is recognized as an impediment to meta-analysis, undermining the validity of pooled effect sizes. [45] A number of statistical studies have shown that ignoring correlation between outcomes leads to biased pooled effect sizes when ORB is present. [24,36] The methodological development and simulation study in this paper aims to demonstrate settings where multivariate network meta-analysis can effectively reduce the impact of ORB by borrowing strength across more fully observed outcomes. The approach we present is especially useful for small to medium-sized networks with a reasonable amount of missing data in at least one outcome, and a moderate to strong correlation between outcomes. Which approach to use must also be driven by subject matter – in the current setting, both efficacy and safety are important considerations in determining optimal treatments for depression in older adults,[46] thus the multivariate approach is sensible.

Results from our simulations (both reported and unreported) demonstrate that MNMA substantially reduces bias to nearly zero under MAR, and also effectively reduce bias under MNAR when the correlation between outcomes is moderate or strong. When the MNMA is applied to networks with complete or MCAR data, there is no loss in efficiency and the same or similar results are obtained as the UNMA approach of Lu and Ades.[2] That the method reduces bias (albeit not to 0) under the MNAR setting is also an important finding for those concerned about the impact that ORB will have on the validity of the pooled effect sizes. The bias reduction is also more profound when correlation between outcomes is stronger. The model accommodates both negatively and positively correlated outcomes via specification of several correlation parameters in the covariance matrix parameters. Overall,

given the availability of freeware OpenBUGS and code presented in the Supplementary Material, it is straightforward to implement the Bayesian MNMA.

There are some limitations of MNMA. First MNMA does not accommodate publication bias. The methods of Mavridis et al [15] should be adapted in future work to accommodate publication bias. Secondly, MNMA can increase bias in estimated effect sizes for fully observed outcomes when the other outcomes are under MNAR, a limitation that has already been addressed in the multivariate pairwise meta-analysis literature. [24,26,47] Similar to those studies, in our simulation study, the MNMA approach introduces a small amount of bias for the fully observed outcome (outcome 1), when the other outcomes are subjected to MNAR missingness. This bias is minimal, though, and is offset by the benefit of bias reduction in other outcomes. Second, it is more computationally complex to employ the multivariate versus univariate method. As previous studies have addressed, [8,29] MNMA requires assumptions about multivariate consistency for which there is no powerful test, as well as estimation of parameters within a potentially large variance-covariance matrix to model both the correlation between outcomes and treatments. In addition, multivariate NMA also suffers from the same limitations as does univariate NMA. A limitation of our simulation work includes that we assumed that the within-study correlations are same by generating aggregate data, primarily to ensure estimability. Although we initially assume within-study correlations are not available, comparisons with the model that accounts for within-study correlation based on individual patient data (IPD) are also desirable. Our approach does not utilize any up/down weighting of multivariate outcomes according to clinical input. As such an approach would be consistent with Bayesian philosophy, it is an area of interest for future study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Petropoulou M, Nikolakopoulou A, Veroniki A-A, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. Journal of clinical epidemiology. 2017;82:20–28. [PubMed: 27864068]
2. Lu G, Ades A. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. Journal of the American Statistical Association. 2006:447–459.
3. Dias S, Welton N, Caldwell D, Ades A. Checking consistency in mixed treatment comparison meta-analysis. Statistics in medicine. 2010;29(7-8):932–944. [PubMed: 20213715]
4. Lumley T Network meta-analysis for indirect treatment comparisons. Statistics in medicine. 2002;21(16):2313–2324. [PubMed: 12210616]
5. Hong H, Chu H, Zhang J, Carlin BP. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. Research synthesis methods. 2016;7(1):6–22. [PubMed: 26536149]
6. Glenny A, Altman D, Song F, et al. Indirect comparisons of competing interventions. NCCHTA; 2005.
7. Caldwell DM, Ades A, Higgins J. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. British Medical Journal. 2005;7521:897.

8. Liu Y, DeSantis SM, Chen Y. Bayesian mixed treatment comparison meta-analysis for correlated outcomes subject to reporting bias. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2018.
9. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. Jama. 2006;295(6):676–680. [PubMed: 16467236]
10. Ghaemi SN, Shirzadi AA, Filkowski M. Publication bias and the pharmaceutical industry: the case of lamotrigine in bipolar disorder. The Medscape Journal of Medicine. 2008;10(9):211. [PubMed: 19008973]
11. Trinquart L, Abbé A, Ravaud P. Impact of reporting bias in network meta-analysis of antidepressant placebo-controlled trials. PLoS One. 2012;7(4):e35219. [PubMed: 22536359]
12. Roest AM, de Jonge P, Williams CD, de Vries YA, Schoevers RA, Turner EH. Reporting bias in clinical trials investigating the efficacy of second-generation antidepressants in the treatment of anxiety disorders: a report of 2 meta-analyses. JAMA psychiatry. 2015;72(5):500–510. [PubMed: 25806940]
13. Bafeta A, Trinquart L, Seror R, Ravaud P. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. BMJ (Clinical research ed). 2012;347:f3675–f3675.
14. Mills EJ, Ioannidis JP, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. Jama. 2012;308(12):1246–1253. [PubMed: 23011714]
15. Mavridis D, Welton NJ, Sutton A, Salanti G. A selection model for accounting for publication bias in a full network meta-analysis. Statistics in Medicine. 2014;30(33):5399–5412.
16. Trinquart L, Chatellier G, Ravaud P. Adjustment for reporting bias in network meta-analysis of antidepressant trials. BMC Medical Research Methodology. 2012;12:150–150. [PubMed: 23016799]
17. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Research Synthesis Methods. 2012;3(2):161–176. [PubMed: 26062088]
18. Mavridis D, Efthimiou O, Leucht S, Salanti G. Publication bias and small-study effects magnified effectiveness of antipsychotics but their relative ranking remained invariant. Journal of clinical epidemiology. 2016;69:161–169. [PubMed: 26210055]
19. Trinquart L, Ioannidis JP, Chatellier G, Ravaud P. A test for reporting bias in trial networks: simulation and case studies. BMC medical research methodology. 2014;14(1):112. [PubMed: 25262204]
20. Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. Statistics in medicine. 2013;32(1):51–66. [PubMed: 22806991]
21. Chan A- W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. Jama. 2004;291(20):2457–2465. [PubMed: 15161896]
22. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS one. 2008;3(8):e3081. [PubMed: 18769481]
23. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. Bmj. 2010;340:c365. [PubMed: 20156912]
24. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in medicine. 2012;31(20):2179–2195. [PubMed: 22532016]
25. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. BMC Medical Research Methodology. 2007;7(1):1. [PubMed: 17217545]
26. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. Statistics in Medicine. 2011;30(20):2481–2498. [PubMed: 21268052]

27. Frosi G, Riley RD, Williamson PR, Kirkham JJ. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. Journal of clinical epidemiology. 2015;68(5):542–550. [PubMed: 25537265]
28. Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bagos P, Salanti G. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. Statistics in Medicine. 2014;33(13):2275–2287. [PubMed: 24918246]
29. Efthimiou O, Mavridis D, Riley RD, Cipriani A, Salanti G. Joint synthesis of multiple correlated outcomes in networks of interventions. Biostatistics. 2015;16(1):84–97. [PubMed: 24992934]
30. Achana FA, Cooper NJ, Bujkiewicz S, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. BMC medical research methodology. 2014;14(1):1. [PubMed: 24383436]
31. Jackson D, Bujkiewicz S, Law M, Riley RD, White I. A matrix-based method of moments for fitting multivariate network meta-analysis models with multiple outcomes and random inconsistency effects. arXiv preprint arXiv:170509112. 2017.
32. Hong H, Carlin B, Chu H, Shamliyan T, Wang S, Kane R. A Bayesian Missing Data Framework for Multiple Continuous Outcome Mixed Treatment Comparisons. AHRQ report (No 13-EHC004-EF). 2013.
33. Dias S, Ades A. Absolute or relative effects? Arm-based synthesis of trial data. Research synthesis methods. 2016;7(1):23–28. [PubMed: 26461457]
34. Thorlund K, Druyts E, Wu P, Balijepalli C, Keohane D, Mills E. Comparative Efficacy and Safety of Selective Serotonin Reuptake Inhibitors and Serotonin-Norepinephrine Reuptake Inhibitors in Older Adults: A Network Meta-Analysis. Journal of the American Geriatrics Society. 2015;63(5):1002–1009. [PubMed: 25945410]
35. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581–592.
36. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Biostatistics. 2008;9(1):172–186. [PubMed: 17626226]
37. Salanti G, Higgins JP, Ades A, Ioannidis JP. Evaluation of networks of randomized trials. Statistical methods in medical research. 2008;17(3):279–301. [PubMed: 17925316]
38. Wei Y, Higgins J. Bayesian multivariate meta-analysis with multiple outcomes. Statistics in medicine. 2013;32(17):2911–2934. [PubMed: 23386217]
39. Bujkiewicz S, Thompson JR, Sutton AJ, et al. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. Statistics in Medicine. 2013;32(22):3926–3943. [PubMed: 23630081]
40. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis Texts in statistical science series. In: Chapman & Hall/CRC, Boca Raton, FL; 2004.
41. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2009;172(4):789–811.
42. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. The Lancet. 2009;373(9665):746–758.
43. DeSantis SM, Zhu H. A Bayesian mixed-treatment comparison meta-analysis of treatments for alcohol dependence and implications for planning future trials. Medical Decision Making. 2014;34(7):899–910. [PubMed: 24935915]
44. Yan J Enjoy the joy of copulas: with a package copula. Journal of Statistical Software. 2007;21(4):1–21.
45. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. bmj. 2014;349:g6501. [PubMed: 25416499]
46. Coupland CA, Dhiman P, Barton G, et al. A study of the safety and harms of antidepressant drugs for older people: a cohort study using a large primary care database. 2011.
47. Riley RD, Sutton AJ, Abrams KR, Lambert PC. Sensitivity analyses allowed more appropriate and reliable meta-analysis conclusions for multiple outcomes when missing data was present. Journal of clinical epidemiology. 2004;57(9):911–924. [PubMed: 15504634]

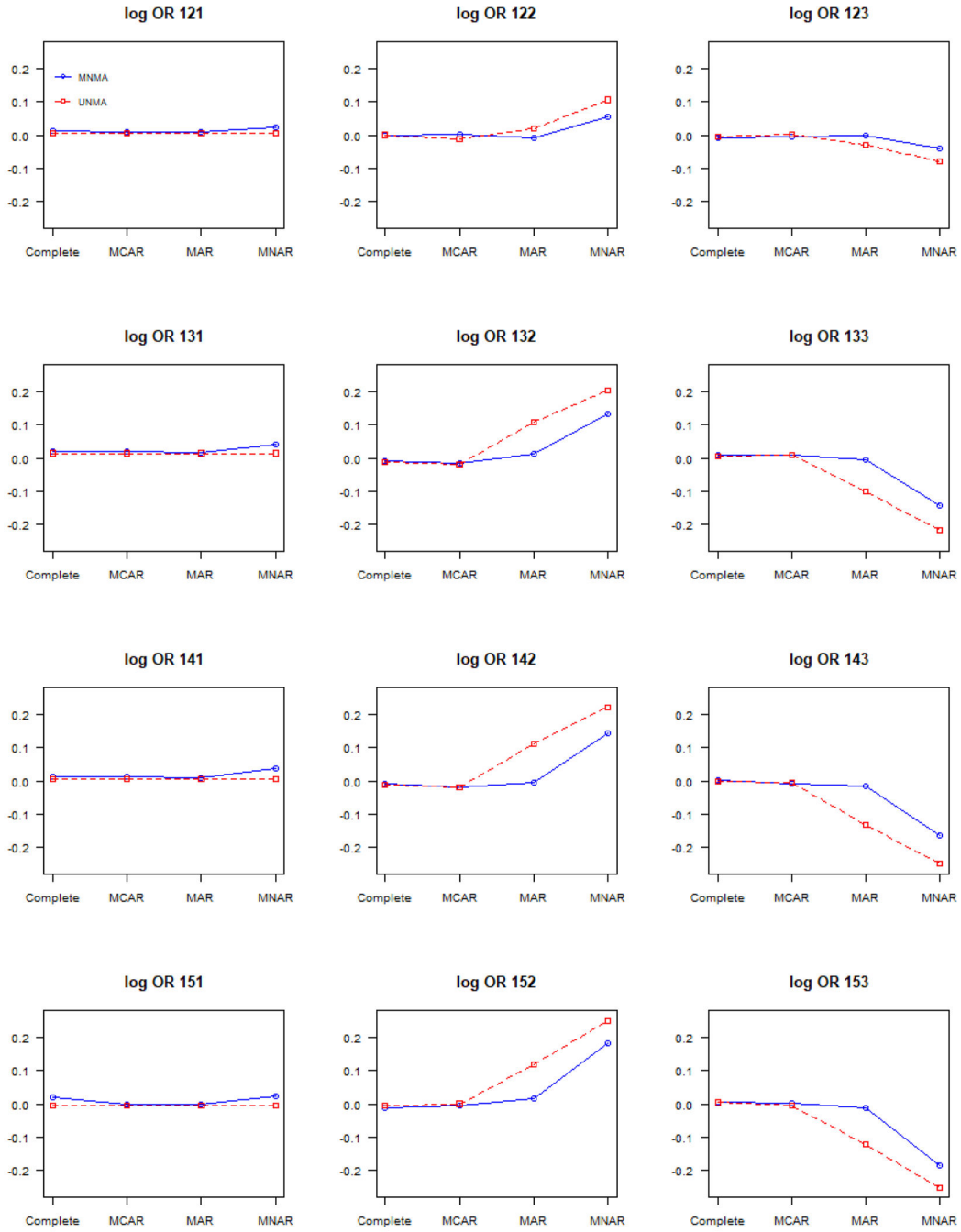

Figure 1.
Bias in estimates of LORs under various missingness scenarios. ($|\rho| = 0.5$, 50% missingness)

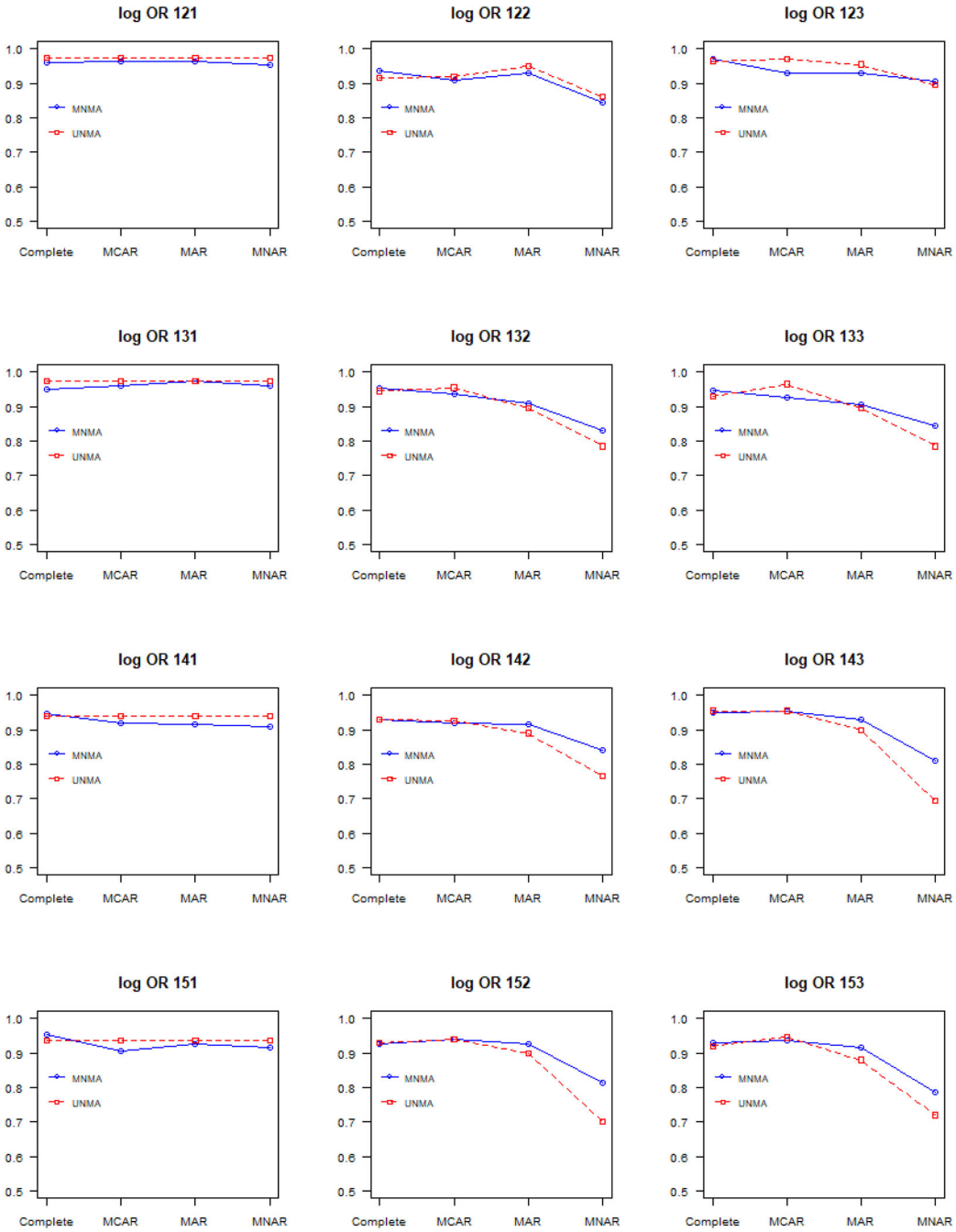

Figure 2.
Coverage probabilities for LORs under various missingness scenarios. ($|\rho| = 0.5$, 50% missing)

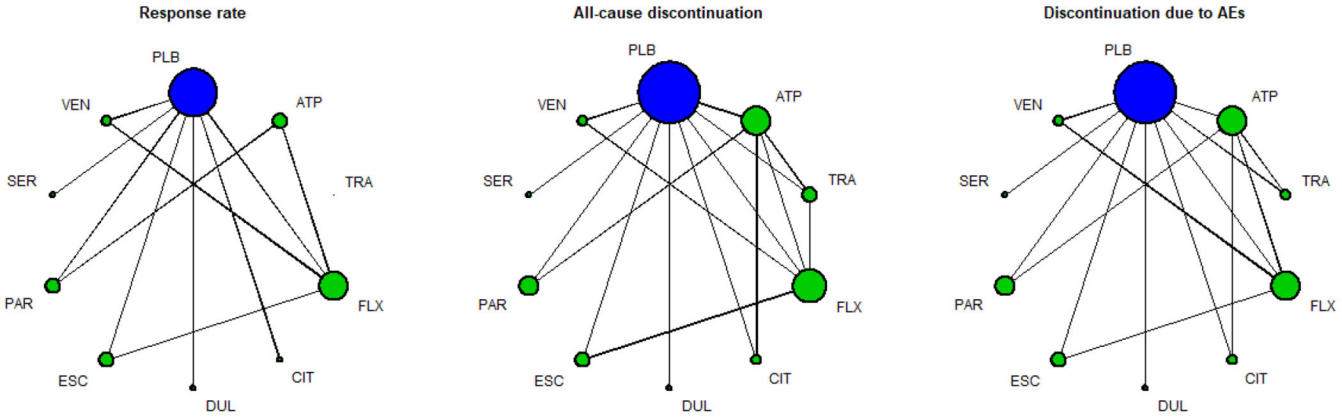

Figure 3.

Network structure of the RCTs for response rate, all-cause discontinuation, and discontinuation due to AEs. PLB=placebo; ATP= Amitriptyline; TRA= Trazodone; FLX=Fluoxetine; CIT=Citalopram; DUL=duloxetine; ESC=Escitalopram; PAR=Paroxetine; SER=Sertraline; VEN=Venlafaxine.

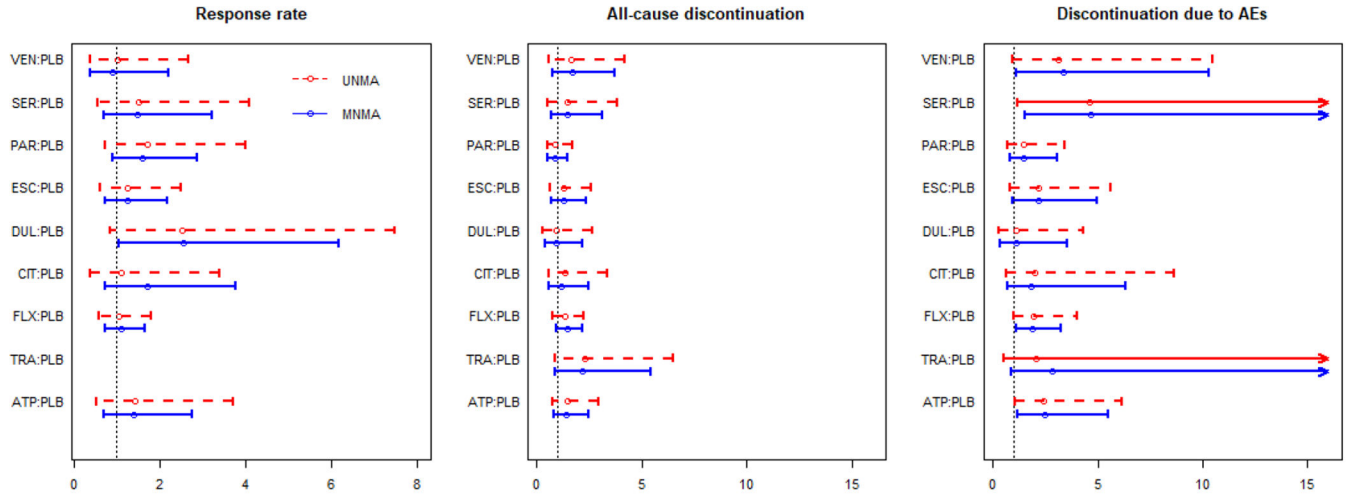

Figure 4.

Posterior odds ratios for response rate, all-cause discontinuation, and discontinuation due to AEs (left to right) for all treatments versus PLB. Dotted and solid lines represent the estimates of the univariate and multivariate NMA, respectively. PLB=placebo; ATP= Amitriptyline; TRA= Trazodone; FLX=Fluoxetine; CIT=Citalopram; DUL=duloxetine; ESC=Escitalopram; PAR=Paroxetine; SER=Sertraline; VEN=Venlafaxine.

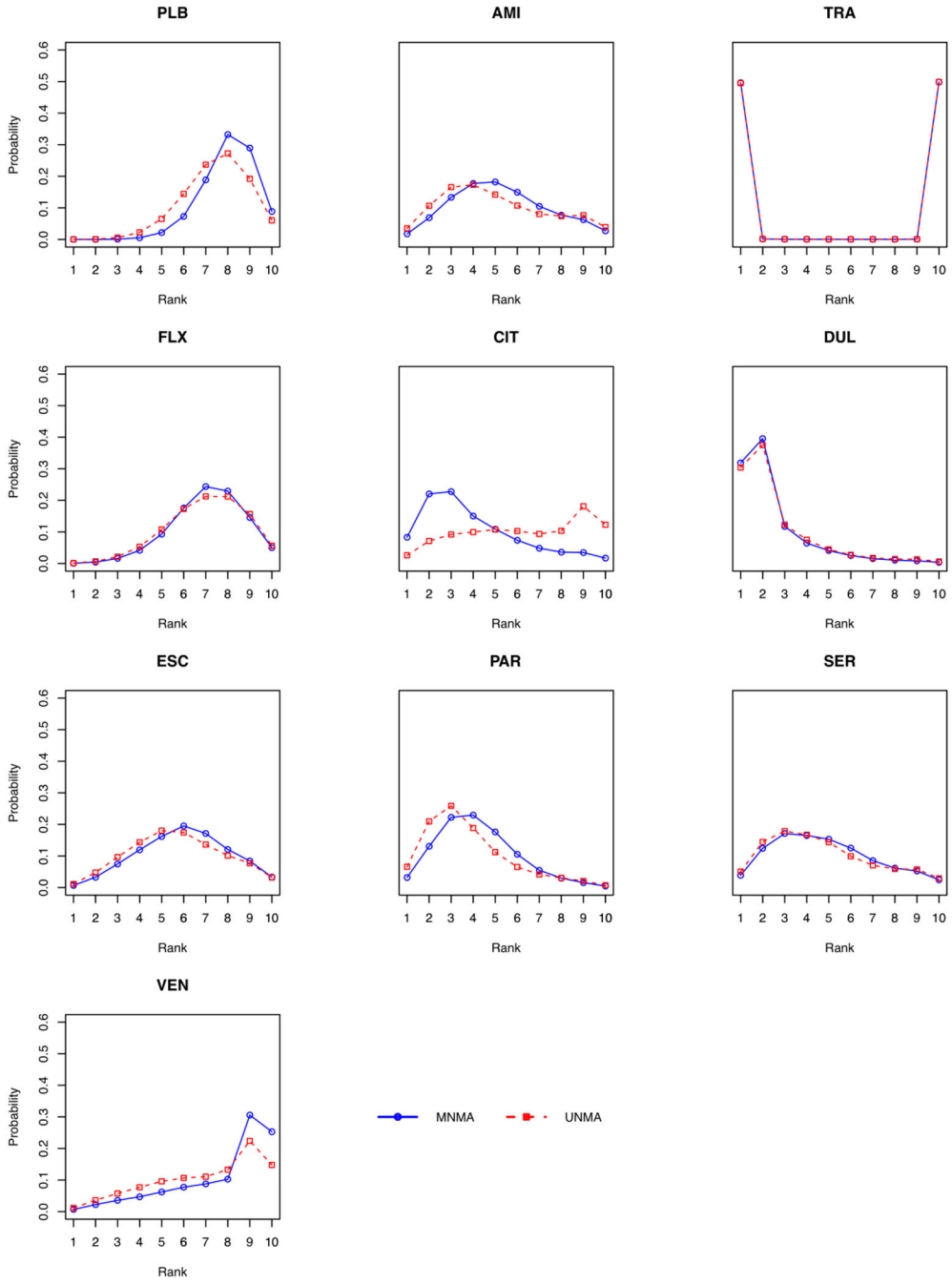

Figure 5.

Ranking distribution for response outcome. The result for diazepam was excluded. PLB=placebo; ATP= Amitriptyline; TRA= Trazodone; FLX=Fluoxetine; CIT=Citalopram; DUL=duloxetine; ESC=Escitalopram; PAR=Paroxetine; SER=Sertraline; VEN=Venlafaxine.

Table 1.

Posterior median estimates and 95% credible intervals for variance of heterogeneity, and posterior mean estimates and 95% credible intervals for correlation parameters including three outcomes for older adults' depression network. Response, all-cause discontinuation, and discontinuation due to AEs outcomes are illustrated as 1,2, and 3, respectively

	τ^2 / ψ_1^2	τ^2 / ψ_2^2	τ^2 / ψ_3^2	ρ_{12}^g	ρ_{13}^g	ρ_{23}^g
UNMA	0.157 [0.004; 1.092]	0.139 [0.001; 0.832]	0.185 [0.001; 1.346]	-	-	-
MNMA	0.095 [0.004; 0.544]	0.076 [0.000; 0.480]	0.106 [0.000; 0.880]	−0.603 [−0.923; −0.155]	−0.641 [−0.930; 0.040]	0.604 [0.050; 0.905]