Statistics and Its Interface Volume 13 (2020) 511–517

Sample size estimation for future studies using Bayesian multivariate network meta-analysis

Stacia M. DeSantis* and Hyunsoo Hwang

Although systematic reviews of randomized clinical trials (RCTs) are considered the pinnacle of evidence-based medicine, RCTs are often designed to reach a desired level of power for a pre-specified effect size, independent of the current body of evidence. Evidence indicates that sample size calculations for a new RCT should be conducted in the context of a systematic review and meta-analysis of the existing body of evidence. This paper presents a framework to estimate sample size and power for a future study, based on a prospective multivariate network meta-analysis (MNMA) of RCTs. The term "multivariate" refers to powering on (potentially) multiple outcomes. Specifically, a Bayesian MNMA is fit to the existing network and 1000 hypothetical trials are designed from the resultant posterior predictive distribution of effect sizes. Thus, the future RCT is designed in the context of the current network of evidence. The approach is applied to a systematic review of pharmacologic treatments for adult acute manic disorder. The analysis suggests that new trials should be designed/powered within the context of either a multivariate or univariate network meta-analysis, where the former is preferred if researchers are interested in multiple primary outcomes, or the network is subject to extensive missing outcomes.

Keywords and phrases: Network meta-analysis, Multivariate meta-analysis, Bayesian, Clinical trials, Sample size.

1. INTRODUCTION

Randomized clinical trials (RCTs) are the foundation of evidence-based medicine [1, 2]. As such, high quality RCTs are often collected and synthesized in the form of a systematic review and meta-analysis [2, 3]. Lower forms of evidence may also be collected and analyzed via systematic reviews and meta-analyses (e.g., case series, observational, and cohort studies). The Cochrane collaboration is a successful network that publishes such systematic reviews that are considered the pinnacle of evidence-based medicine, with medical decision-making relying on these results [1, 4].

*Corresponding author.

In practice, it is not common for a single study to reach a decisive result. Trials may be underpowered, subject to missing outcome data, or even stopped early. Often, subsequent trials or studies are designed to reach a desired level of power in order to detect a pre-specified effect size through sample size adjustment. However, new sample size calculations usually do not take into account the existing body of evidence, nor the impact the new trial will have on the existing body of evidence. In the interest of efficiency, harm-reduction, and cost-efficiency, researchers should consider the influence their individual study will have on future systematic reviews, as well as the influence their findings will have on medical decision making, given the current available body of evidence [5–8].

A few meta-analytic frameworks have been developed that make this suggestion. Sutton et al. [5] proposed estimating sample size using a hybrid framework. Specifically, they used both fixed- and random-effects meta-analysis to estimate a Bayesian posterior predictive distribution (PPD), sample an effect size for a new trial, and calculate power using simulations subsequently performed under a frequentist framework. DeSantis and Zhu [6] extended this approach under a fully Bayesian setting, to network meta-analysis (NMA) [9, 10]. Roloff et al. [7] introduced the conditional probability (CP) method, which unlike the above, does not require a simulation procedure; the CP is defined as the probability that effect estimates from an updated meta-analysis will exceed a pre-specified effect size, given the pooled results of existing meta-analyses. Nikolakopoulou et al. [8] extended the CP approach to the context of NMA, assuming consistency in the network and common heterogeneity in old and new studies given multiple comparisons. They investigated the power of a new meta-analysis in relation to both the number of studies, and the comparison types in the network (direct vs indirect comparisons of treatments), given a fixed sample size for the current study.

These approaches should all serve as a guide on how to estimate sample size for a new trial in the context of a prospective meta-analysis. In scenarios where a systematic review yields a large proportion of missing outcomes, for example due to outcome reporting bias (ORB), trial design in the context of multivariate meta-analysis may be preferable. The issue of ORB is well known to afflict systematic reviews, and its effects on bias in both pairwise and network meta-analysis have therefore been well-studied [11–19]. Briefly,

ORB may occur when an outcome of interest for a given study was analyzed but not reported, perhaps due to lack of significance or unfavorable directionality [11, 12, 17, 18]. When ORB is present within a systematic review, a prospective predictive sample size based on the PPD from univariate meta-analysis, may not be optimal. To this end, sample size could be estimated from multivariate meta-analysis or multivariate network meta-analysis (MNMA); these tools have been proposed to accurately pool effect sizes in the context of ORB, usually by borrowing power across correlated outcomes [11–19]. Those approaches that operate within a Bayesian framework are easy to implement in practice, are well integrated into commercial and freely available software, and naturally lend themselves to calculating posterior predictive distributions [5, 6].

The objective of this paper is to present the framework to estimate sample size and power for a future RCT based on a prospective MNMA. The proposed framework is an extension of the methods proposed by Sutton et al. [5] and DeSantis and Zhu [6]. Using a previously published network of 12 pharmacological treatments for acute manic disorder [20], sample sizes for hypothetical 2-arm trials are simulated under NMA and MNMA for fixed sample sizes, and power is compared versus had the trials been designed in isolation of the networks.

2. METHODS

2.1 Multivariate network meta-analysis model

This Section introduces the MNMA model using the method of Efthimiou et al. [15] and Hwang and DeSantis [17] though any MNMA framework could likely be used so long as the PPD can be estimated. Consider N_T total treatments in a network with a maximum of 3 outcomes. The consistency equations of Lu and Ades [9] for each outcome imply 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, i.e.,

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

In a random effects multivariate network meta-analysis, the parameter of interest, $\beta_{i,(B:C)l}$ is a pooled log odds ratio of treatment C relative to treatment B for outcome l in the ith study. For 2-arm studies reporting 3 outcomes, Hwang and DeSantis [17] the MNMA model can be written, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\nu}$,

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

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, $\boldsymbol{\beta}$ is an $(N_T - 1) \times 3$-dimensional vector of basic parameters, $\boldsymbol{\nu}$ is the combined vector of random errors and additional variations due to heterogeneity, with $\boldsymbol{\nu} \sim \mathrm{N}(0,\ \boldsymbol{\Sigma})$. In this model, the heterogeneity is assumed to be constant between different comparisons. The MNMA approach uses the design matrix $\boldsymbol{X}$ to appropriately map the observed treatment comparisons, using basic parameters, and the variance-covariance matrix, Σ_{i}. To reduce the burden of parameter estimation, Efhthimiou et al. [15] reduces the complexity of the variance-covariance matrix using a homogenous variance assumption. The simplified variance-covariance matrix for 3-arms and 3 outcomes that compares treatments A, B, and C is given as,

(1)

$$\Sigma_i = \begin{pmatrix} \upsilon_{i,AB,1} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho^g_{12}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AB,2}} & \upsilon_{i,AB,2} & \cdot & \cdot & \cdot & \cdot \\ \rho^g_{13}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AB,3}} & \rho^g_{23}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AB,3}} & \upsilon_{i,AB,3} & \cdot & \cdot & \cdot \\ \frac{1}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,1}} & \frac{\rho^g_{12}}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,1}} & \frac{\rho^g_{13}}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,1}} & \upsilon_{i,AC,1} & \cdot & \cdot \\ \frac{\rho^g_{12}}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,2}} & \frac{1}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,2}} & \frac{\rho^g_{23}}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,2}} & \rho^g_{12}\sqrt{\upsilon_{i,AC,1}\upsilon_{i,AC,2}} & \upsilon_{i,AC,2} & \cdot \\ \frac{\rho^g_{13}}{2}\sqrt{\upsilon_{i,AB,1}\upsilon_{i,AC,3}} & \frac{\rho^g_{23}}{2}\sqrt{\upsilon_{i,AB,2}\upsilon_{i,AC,3}} & \frac{1}{2}\sqrt{\upsilon_{i,AB,3}\upsilon_{i,AC,3}} & \rho^g_{13}\sqrt{\upsilon_{i,AC,1}\upsilon_{i,AC,3}} & \rho^g_{23}\sqrt{\upsilon_{i,AC,2}\upsilon_{i,AC,3}} & \upsilon_{i,AC,3} \end{pmatrix},$$

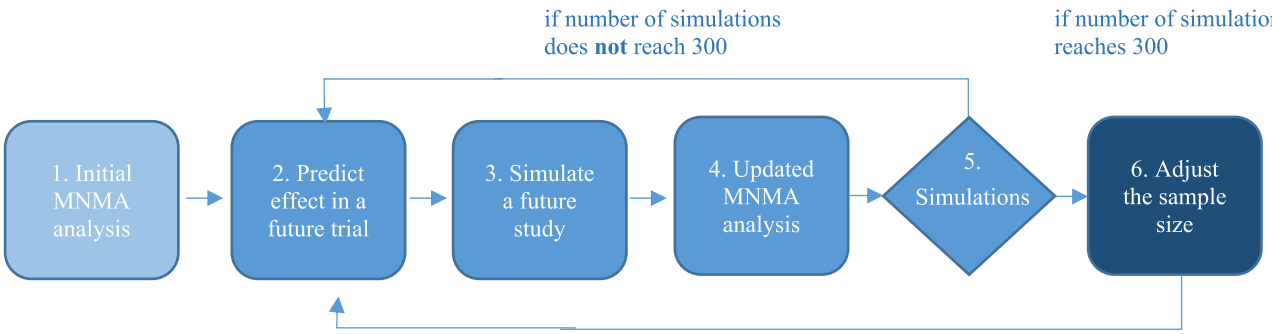

Figure 1. Diagram representing the sample size estimation process in a context of the prospective meta-analysis.

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. Thus to ensure the variance-covariance matrices are positive definite, the Cholesky decomposition is used. This is the MNMA used in the analysis described below. More technical detail can be found in the references [15, 17].

2.2 MNMA sample size and power estimation framework

Posterior predictive distributions are generated from a MNMA fit using Markov chain Monte Carlo (MCMC) sampling, and operating characteristics (e.g. power and sample size) for a new trial are based on whether the pooled evidence incorporating the new trial would result in a conclusive or decisive finding [5, 6]. The term "decisive" is used here to imply Bayesian significance, or the exclusion of the null effect size by the 95% credible interval (CI).

The simple framework to estimate sample size and power based on a prospective MNMA is depicted in Figure 1, and the proposed algorithm is given in 6 steps:

1. Run a Bayesian MNMA on the existing body of evidence (described for the below data example using [17], noting one can use any of the methods presented in [14–18] could be used if outcomes were commensurate with those methods).
2. Predict the effect size in a future trial. That is, within the MCMC sampler, sample an estimate, $\theta_{XY;l(\text{new})}$, from the PPD given by,

$$\begin{pmatrix}\theta_{XY;1(\text{new})}\\ \theta_{XY;2(\text{new})}\end{pmatrix} \sim N\left(\begin{pmatrix}\beta_{XY;1}\\ \beta_{XY;2}\end{pmatrix}, \mathbf{\Phi} = \begin{pmatrix}\psi_1^2 & \rho^g\psi_1\psi_2\\ \rho^g\psi_1\psi_2 & \psi_2^2\end{pmatrix}\right),$$

   where $\beta_{XY;l}$ is the true effect size for outcome l (for $l = 1, 2$ in this setting), $\mathbf{\Phi}$ is the covariance matrix where ψ represents the variation due to heterogeneity between studies, ρ^g is the global correlation parameter, and X & Y are two treatments of interest for future study (where X is regarded as the baseline treatment). As previously studied, the variation due to heterogeneity, ψ, is not exactly equivalent to the between-study variance in the standard hierarchical model although it can be comparable [14, 15, 17]. In the Bayesian setting, uncertainty in β, ρ^g, and ψ would have propagated through the Bayesian MNMA model to the posterior predictive distribution of $\theta_{XY;l(\text{new})}$ for a new study. The PPD is therefore more overdispersed than the posterior distribution of $\theta_{XY;l}$, and is probably the best "guess" for designing a future trial.
3. Given multiple samples from the PPD, $\theta_{XY;l(\text{new})}$, simulate a new study with a complete outcome vector. Assuming the baseline (X, e.g., placebo) event rate $P_{X;l(new)}$ is known, calculate the event rate within treatment Y, $P_{Y;l(new)}$, for each outcome as follows,

$$P_{Y;l(new)} = \frac{((P_{X;l(new)}/(1-P_{X;l(new)})) \times e^{\theta_{XY;l(\text{new})}})}{(1+(P_{X;l(new)}/(1-P_{X;l(new)})) \times e^{\theta_{XY;l(\text{new})}})}.$$

   In practice, $P_{X;l(new)}$ are usually unknown but $P_{X;l(new)}$ can be estimated from existing trial data or by eliciting expert opinion. Using the estimates of $P_{X;l(new)}$ and $P_{Y;l(new)}$, generate a new binomial dataset according to a fixed set of sample sizes. Employ a Gaussian Copula distribution to generate correlated binomial outcomes (if powering on multiple outcomes is desired), using a previously sampled global correlation coefficient, ρ^g [18].
4. Run an updated MNMA. That is, combine the new simulated study to the existing network and run a new MNMA to update the posterior distribution (e.g., posterior mean and 95% credible intervals) of $\beta_{XY;l(updated)}$.
5. Run B simulations for the user-determined sample sizes. The procedure of sampling a predicted effect, $\theta_{XY;l(\text{new})}$, simulating a future study, and running an updated MNMA with the additional study, (i.e. steps 2–4), is repeated B times for each user-defined fixed sample sizes. The estimated proportion of times over B simulations that the estimated effect size $\beta_{XY;l(updated)}$ is decisive for each outcome in MNMA (i.e., proportion

of B total times the 95% CI excludes 0 for log odds ratio) is the proposed power of the study for each outcome. A lower bound for B could be 300 as MNMA can have long computation time depending on the constitution of the network.

6. Repeat steps 2–5 for each sample size until the desired power is obtained. In the following study, sample sizes are fixed at clinically feasibly sizes of (30, 50, 100, 150, 200, 300, and 500).

2.3 Data Description and illustration

The application is a published network of interventions for acute manic disorder [20]. The network comprises 68 randomized controlled trials that compare the effectiveness and tolerability of 13 active treatments and a placebo. Treatments include olanzapine, paliperidone, quetiapine, divalproex, aripiprazole, carbamazepine, haloperidol, ziprasidone, asenapine, lithium, lamotrigine, topiramate, and gabapentin [20]. The majority of studies (50) are 2 arm trials while 18 are 3 arm trials. A total of 16,073 patients are included in the network. The outcomes for each study are response and dropout. Response is defined as the proportion of patients whose mania symptoms were reduced at least 50% from baseline. Dropout is defined as the proportion of patients leaving the study for any reason. Both outcomes are reported as binary and are parametrized as log odds ratios (LOR) for the purpose of meta-analysis. While one study did not report dropout, 18 studies did not report a binary response. Given the unusually high proportion of non-report of the primary outcome (reduction in symptoms), it is of interest to design the future study for a bivariate outcome, such that dropout may inform the sample size in addition to the primary outcome.

In a previously conducted univariate network meta-analysis (UNMA) of these data [20], most of the active treatments were shown to be superior to the placebo, with the exception of lamotrigine and topiramate. Those results are also consistent with the bivariate NMA reported by Efthimiou et al. [15] Following closely this work, the current analysis designs two different hypothetical 2-arm trials comparing the active treatments that were previously shown superior to placebo. The first hypothetical trial is a comparison between olanzapine and lithium. Olanzapine is often considered a first-line choice with paliperidone due to high efficacy, but results from univariate [20] and multivariate [15] NMA showed it was not significantly superior to lithium (which has been used in the treatment of acute bipolar mania for over 50 years and serves as the traditional treatment option [21]). Although lithium is underused due to relatively slow response and poor tolerability, it is still regarded as a gold-standard comparator for other newer agents [21, 22]. Thus a new trial comparing olanzapine vs lithium could be of interest for researchers in order to provide a rationale for choosing between either treatment as a first-line therapy. The second hypothetical trial is a comparison between carbamazepine and divalproex. Both treatments are anticonvulsants that have been shown to be equally effective in treating mixed and classic mania [21]. Although they are potentially equivalent in their effects, the effectiveness of one over the other has not been established. Thus, the below reports on hypothetical trial designs comparing these treatments, based on the existing multivariate network of evidence.

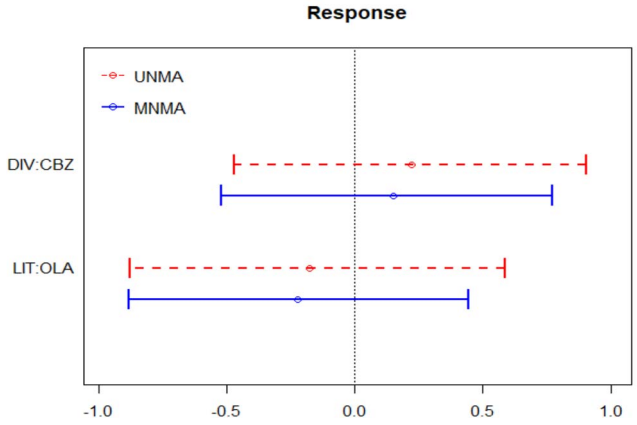

Figure 2. Posterior mean log odds ratio and 95% credible interval for UNMA (red and dotted) and multivariate NMA (blue and solid) comparisons. A log odds ratio <0 favors the second treatment and log odds ratio excluding zero indicates a significant difference. LIT, OLA, DIV, and CBZ denote lithium, olanzapine, divalproex, and carbamazepine, respectively.

3. RESULTS

Figure 2 presents the results from applying UNMA and MNMA using the approaches and code presented in the Methods Section adapted from [15, 17]. The univariate analysis only considers treatment response (reduction in mania symptoms greater than 50%). The multivariate analysis considers both treatment response and discontinuation. The pooled posterior mean ORs (95% CIs) for lithium vs olanzapine are −0.18 (−0.88, 0.59) and −0.22 (−0.89, 0.44) from univariate and multivariate NMA, respectively. The pooled LORs for the divalproex vs carbamazepine comparison are 0.22 (−0.47, 0.90) and 0.15 (−0.52, 0.77), respectively. The results from both approaches are directionally similar, which is expected, but there is some difference in the actual posterior mean and width of the 95% credible intervals. The forest plot in Figure 2 demonstrates that the multivariate analysis improves the precision of estimates by narrowing the CIs, consistent with prior reports [15, 17]. Furthermore, point estimate changes from fitting the MNMA may be due to a reduction in reporting bias achieved via borrowing power from the correlated (secondary) outcome that is measured in every trial except for one. Since neither 95% CI excludes

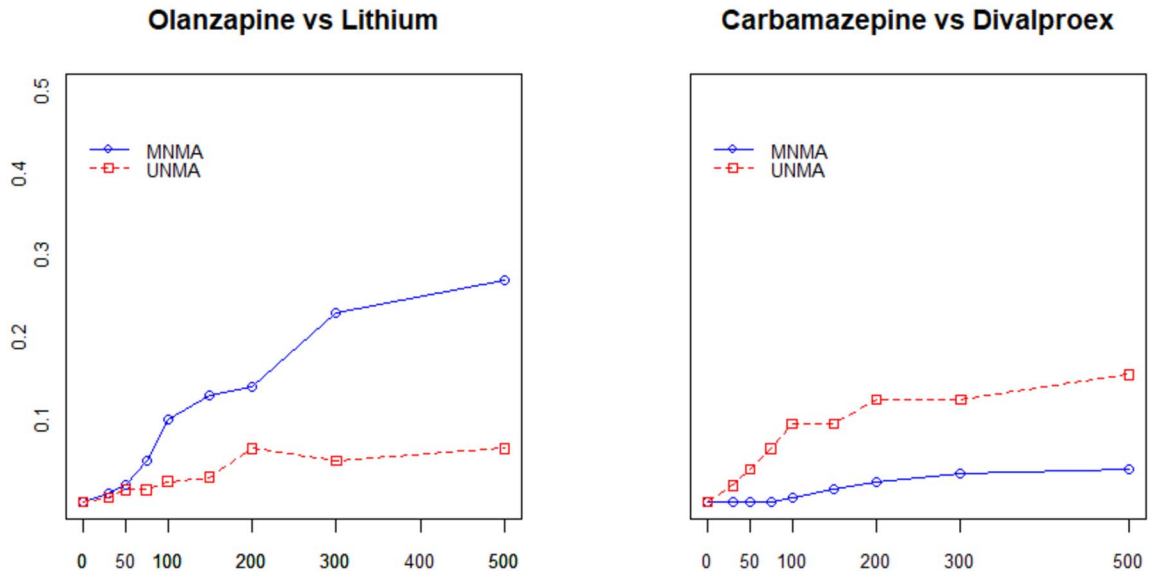

Figure 3. Power (y-axis) and sample size curves for studies of Olanzapine vs Lithium and Carbamazepine vs Divalproex designed under a new univariate and multivariate network meta-analysis.

	PLA	OLA	PAL	QUE	DIV	ARI	CBZ	HAL	ZIP	ASE	LIT	LAM	TOP
PLA		1.00	1.00	1.00	1.00	1.00	0.93	1.00	0.87	0.79	1.00	0.28	0.03
OLA	1.00		0.00	0.00	0.00	0.00	0.03	0.00	0.76	0.15	0.25	0.76	0.98
PAL	1.00	0.00		0.00	0.02	0.00	0.00	0.00	0.70	0.07	0.17	0.73	0.99
QUE	1.00	0.00	0.00		0.00	0.00	0.01	0.00	0.49	0.10	0.00	0.77	0.98
DIV	1.00	0.00	0.00	0.00		0.00	0.09	0.03	0.22	0.05	0.01	0.58	0.88
ARI	1.00	0.00	0.00	0.00	0.00		0.00	0.00	0.43	0.11	0.03	0.60	0.91
CBZ	0.95	0.08	0.12	0.17	0.14	0.14		0.02	0.52	0.15	0.13	0.64	0.90
HAL	1.00	0.00	0.00	0.01	0.03	0.01	0.05		0.80	0.18	0.27	0.71	0.94
ZIP	0.69	0.83	0.61	0.52	0.44	0.50	0.63	0.88		0.17	0.03	0.33	0.79
ASE	0.60	0.10	0.16	0.14	0.13	0.14	0.31	0.21	0.19		0.07	0.54	0.73
LIT	1.00	0.07	0.04	0.01	0.00	0.00	0.27	0.14	0.18	0.09		0.62	0.91
LAM	0.46	0.53	0.50	0.49	0.49	0.37	0.56	0.53	0.44	0.45	0.41		0.46
TOP	0.23	0.89	0.90	0.91	0.89	0.92	0.91	0.95	0.62	0.78	0.80	0.56	

Figure 4. Power for the design of a new trial with sample size 1000. The upper triangle illustrates power based on updated multivariate NMA and the lower triangle illustrates power based on updated UNMA. Abbreviations are: (ARI, aripiprazole; ASE, asenapine; CBZ, carbamazepine; VAL, divalproex; HAL, haloperidol; LAM, lamotrigine; LIT, lithium; OLZ, olanzapine; PBO, placebo; QTP, quetiapine; PAL, paliperidone; TOP, topiramate; ZIP, ziprasidone).

zero, it is straightforward to conclude from both UNMA and MNMA that there is no significant difference in treatments.

The power estimation simulation for a new trial, as detailed in the Methods, is applied to the two hypothetical comparisons. Figure 3 depicts the power at a fixed sample sizes of 2-arm trials resulting from 300 simulations of trials for the treatment comparisons in Figure 2. A new trial of these treatments designed in the context of MNMA would have very low power (less than 0.3 even at $N = 500$ per group), which is consistent with posterior mean observations from Figure 2. Further, Figure 3 shows that the power estimates from a hypothetical trial of carbamazepine vs divalproex designed in the context of MNMA are consistently lower than those designed in the context of UNMA (Figure 3, right side plot). This will not be true as a general rule, but is true for the current study since the inclusion of the second outcome actually decreases the effect size of the primary outcome (reduction in mania symptoms) for this treatment comparison. Both the UMNA and MNMA power estimates for a fixed sample size are also lower than if the trial were to be designed in isolation of the network, or if the trial were to be designed based on the posterior mean effect size (which is underdispersed, data not shown).

Figure 4 considers all hypothetical 2-arm future trials of existing treatments for mania just for completeness. Each 2-arm treatment comparison is represented by row and column, and each box entry is the simulated power resulting from enrolling $N = 2000$ patients (1000 per group), using

the simulation paradigm outlined in the Methods. The upper triangle displays the power resulting from designing the trial via a multivariate approach while the lower triangle displays the power resulting from designing the trial via a univariate approach. It is notable that there is no clear superiority of either method – the benefit of designing the trial under a multivariate vs univariate approach truly depends on whether there is likely to be outcome reporting bias. Figure 4 shows that most trials that compare active treatment to placebo (row 1, column 1) would be highly powered, as expected. Further, trials of lamotrigine, and topiramate would be highly powered. However, hypothetical trials of active treatments haloperidol, carbamazepine, olanzapine, paliperidone, and aripiprazole would add little to the network of evidence. When planning a trial, clinicians could use Figure 4, whether under UNMA or MNMA, to determine whether the treatments they are interested in comparing would result in sufficient power for a fixed set of operating characteristics, in light of the current body of evidence.

4. DISCUSSION

The application of network meta-analysis has become standard practice in systematic reveiws of RCTs[9, 10, 15, 17, 18]. Although power and sample size estimation for a newly proposed clinical trial in the context of an updated NMA has been proposed and is an appropriate way to plan a new study, it has been shown that pooled estimates from NMA could also be biased or inefficient due to the presence of outcome reporting bias [14, 15, 17, 18]. As a result, new trials designed from a potentially biased network may lead to either over- or under-estimated power calculations for a given sample size.

This paper presents an easy-to-adopt approach for estimating power and sample size based on a prospective multivariate network meta-analysis [15, 17]. The software used in this paper is publicly available and downloadable from both the original paper and from GitHub [17, 25]. We primarily advocate the use of MNMA when multiple outcomes are of interest in designing an RCT (for example, the joint outcome of efficacy and discontinuation, which are often analyzed in mental health RCTs), or when primary outcomes of interest are unreported for a large proportion of the studies in the existing network.

The current MNMA framework for trial planning utilizes previously proposed Bayesian meta-analytic procedures [5, 6, 15, 17]. The approach offers benefits over the previously proposed conditional probability methods [7, 8] – adapting the CP method in the context of a MNMA would be computationally challenging due to correlation between multiple outcomes and the presence of multiple (>2) arms within a study.

This study additionally examines how missingness of trial data, likely due to ORB, impacts power and sample size estimation in a prospective manner using an example in acute mania. In the presences of missing outcomes, it has already been well-studied that the multivariate approach results in less biased and more accurate pooled meta-analytic effect sizes; it is therefore also reasonable to assume it results in different prospective power estimates than the univariate approach. In the acute mania example, hypothetical trials of olanzapine vs lithium and carbamazepine vs divalproex, the difference in posterior mean ORs (point estimates) between the univariate and multivariate NMA, led to considerable changes in power estimate for a new trial.

There are several limitations of the current study. First, the approach outlined in this paper is not applicable to novel treatments (for which effect sizes versus existing treatments would be completely unknown). However, if a novel treatment were under study (and not part of the existing network), the power of a future trial could still be estimated in the context of the network. That is, in designing a study of a novel treatment, one could still utilize the existing information for the comparator treatment of interest. Secondly, estimating sample size based on prospective MNMA versus UNMA [6] is computationally more challenging and time consuming, which could limit its utility especially for large networks. However, running simulations in parallel on a computer cluster is straightforward, and decreases computation time. Further, our recently published software for Bayesian MNMA that accommodates any number of arms and outcomes could alleviate the computational or coding challenge for applied researchers attempting to fit this complex model [17, 25].

ACKNOWLEDGMENTS

We acknowledge NIH/NIMH grant number 110110. We also acknowledge Dr. Derek Brown for his editing contributions.

Received 31 October 2019

REFERENCES

[1] Higgins, J. P. T., Green, S. (editors) (2011). *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available from www.handbook.cochrane.org.

[2] Wandel, S., Roychoudhury, S. (2016). Designing and analysing clinical trials in mental health: an evidence synthesis approach. *Evidence Based Mental Health* **19**(4) 114–117.

[3] Liberati, A., Altman, D. G., Tetzlaff, J., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS medicine* **6**(7) e1000100.

[4] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine* **6**(7) e1000097.

[5] Sutton, A. J., Cooper, N. J., Jones, D. R., Lambert, P. C., Thompson, J. R., Abrams, K. R. (2007). Evidence-based sample size calculations based upon updated meta-analysis. *Statistics in medicine* **26**(12) 2479–2500. MR2364400

[6] DeSantis, S. M., Zhu, H. (2014). A Bayesian mixed-treatment comparison meta-analysis of treatments for alcohol dependence and implications for planning future trials. *Medical Decision Making* **34**(7) 899–910.
[7] Roloff, V., Higgins, J., Sutton, A. J. (2013). Planning future studies based on the conditional power of a meta-analysis. *Statistics in medicine* **32**(1) 11–24. MR3017880
[8] Nikolakopoulou, A., Mavridis, D., Salanti, G. (2014). Using conditional power of network meta-analysis (NMA) to inform the design of future clinical trials. *Biometrical Journal* **56**(6) 973–990. MR3270105
[9] Lu, G., Ades, A. (2006). Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *Journal of the American Statistical Association* 447–459. MR2256166
[10] Dias, S., Welton, N., Caldwell, D., Ades, A. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine* **29** 932–944. MR2752057
[11] Kirkham, J. J., Dwan, K. M., Altman, D. G., Gamble, C., Dodd, S., Smyth, R., Williamson, P. R. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *Bmj* **340** c365.
[12] Jackson, D., Riley, R., White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in medicine* **30** 2481–2498. MR2843472
[13] Luo, S., Chen, Y., Su, X., Chu, H. (2014). mmeta: An R Package for Multivariate Meta-Analysis. Department of Biostatistics, Epidemiology and Informatics. 10.
[14] Efthimiou, O., Mavridis, D., Cipriani, A., Leucht, S., Bagos, P., Salanti, G. (2014). An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Statistics in medicine* **33** 2275–2287. MR3256692
[15] Efthimiou, O., Mavridis, D., Riley, R. D., Cipriani, A., Salanti, G. (2015). Joint synthesis of multiple correlated outcomes in networks of interventions. *Biostatistics* **16** 84–97. MR3365413
[16] Achana, F. A., Cooper, N. J., Bujkiewicz, S., Hubbard, S. J., Kendrick, D., Jones, D. R., Sutton, A. J. (2014). Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Medical Research Methodology* **14** 1.
[17] Hwang, H, DeSantis, S. M. (2018). Multivariate network meta-analysis to mitigate the effects of outcome reporting bias. *Statistics in Medicine* **37**(22) 3254–3266. MR3853281
[18] Liu, Y., DeSantis, S. M., Chen, Y. (2018). Bayesian network meta-analysis with correlated outcomes subject to reporting bias, *Journal of The Royal Statistical Society Series C* **67**(1) 127–144. MR3758758
[19] Lin, L., Chu, H. (2018). Bayesian multivariate meta-analysis of multiple factors. *Research Synthesis Methods*. https://doi.org/10.1002/jrsm.1293.
[20] Cipriani, A., Barbui, C., Salanti, G., et al. (2011). Comparative effectiveness and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet* **378** 1306–1315.
[21] Vieta, E, Sanchez-Moreno, J. (2008). Acute and long-term treatment of mania. *Dialogues in clinical neuroscience* **10**(2) 165.
[22] Vieta, E., Goigolea, J. (2010). Treatment guidelines for acute mania. *Annals of General Psychiatry* **9**(Suppl 1) S62.
[23] Lee, A. W. (2014). Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *Journal of clinical epidemiology* **67**(2) 138–143.
[24] Lee, A. W. (2016). Use of network meta-analysis in systematic reviews: a survey of authors. *Systematic reviews* **5**(1) 8.
[25] Hwang, H., DeSantis, S. M. (2019). Multivariate network meta-analysis. GitHub. https://doi.org/10.5281/zenodo.2581561.

Stacia M. DeSantis
University of Texas Health Science Center at Houston
Department of Biostatistics and Data Science
1200 Pressler St
Houston, TX 77035
USA
E-mail address: stacia.m.desantis@uth.tmc.edu

Hyunsoo Hwang
University of Texas MD Anderson Cancer Center
Department of Biostatistics
Houston, TX
USA
E-mail address: hhwang@mdanderson.org