Handling missing continuous outcome data in a Bayesian network meta-analysis

Danila Azzolina [1], *Ileana Baldi* [1], *Clara Minto* [1], *Daniele Bottigliengo* [1], *Giulia Lorenzoni* [1], *Dario Gregori* [1],

(1) Unit of Biostatistics, Epidemiology and Public Health, Dept. of Cardiac Thoracic Vascular Sciences and Public Health, University of Padova, Italy

CORRESPONDING AUTHOR: *Prof. Dario Gregori, Unit of Biostatistics, Epidemiology and Public Health, Dept. of Cardiac Thoracic Vascular Sciences and Public Health, Via Loredan, 18, 35121 Padova – Italy Phone: +39 049 8275384, Fax: +39 027 00445089, Email: dario.gregori@unipd.it*

ABSTRACT

Background: A Bayesian network meta-analysis (NMA) model is a statistical method aimed at estimating the relative effects of multiple interventions against the same disease. The method has recently gained prominence, leading to the synthesis of the evidence regarding rank probabilities for each treatment. In several cases, an NMA is performed excluding incomplete data of studies retrieved through a systematic review, resulting in a loss of precision and power.
Methods: There are several methods for handling missing or incomplete data in an NMA framework, especially for continuous outcomes. In certain cases, only baseline and follow-up measurements are available; in this framework, to obtain data regarding mean changes, it is necessary to consider the pre-post study correlation. In this context, in a Bayesian setting, several authors suggest imputation strategies for pre-post correlation. In other cases, a variability measure associated with a mean change score might be unavailable. Different imputation methods have been suggested, such as those based on maximum standard deviation imputation. The purpose of this study is to verify the robustness of Bayesian NMA models concerning different imputation strategies through simulations.
Results: Simulation results show that the bias is notably small for every scenario, confirming that rankings provided by models are robust concerning different imputation methods in several heterogeneity-correlation settings.
Conclusions: This NMA method seems to be more robust to missing data imputation when data reported in different studies are generated in a low-heterogeneity scenario. The NMA method seems to be more robust to missing value imputation if the expectation of the prior distribution, defined on the heterogeneity parameter, approaches the true value of the variability across studies.

Key words: NMA, missing imputation, Bayesian, continuous data

INTRODUCTION

Network meta-analysis (NMA) is an increasingly popular statistical method used to perform a simultaneous comparison in the effectiveness of treatments against the same disease [1, 2]. The possibility to summarize evidence from studies reporting direct treatment comparisons and to infer indirect effects between interventions that have not been directly compared in head-to-head studies makes this method very attractive for clinical researchers [3].

NMA's advantage over classical pairwise meta-analysis lies in gathering direct and indirect evidence to provide comparisons between the relative effectiveness of all treatments included in the network [4]. In this framework, calculations of the probability of one treatment being the best or worst for a specific outcome through the final ranking of available treatments against the same disease are usually employed to facilitate the interpretation of the results [5, 6]. NMA models are available for a broad range of underlying data and summary effect measures, being implemented in both frequentist and Bayesian frameworks [7].

Among studies identified through a systematic literature review, only those reporting complete data on a treatment effect and its uncertainty usually contribute to an NMA model. In the literature, it is assessed that, for quantitative synthesis of evidence, the continuous outcome is more difficult to synthesize in an NMA than a binary endpoint for several reasons. For example, the research articles may not report all the information required to perform a meta-analysis [8]. The most common method to analyze this kind of outcome in meta-analysis is to consider the mean change from baseline, including also an associated variability measure (standard deviation SD) [9]. However, in a different situation [10], no or incomplete variability measures have been reported in clinical trial publications [11]. In several cases, an NMA performed on a continuous endpoint involves the exclusion of the trial from the NMA due to the incompleteness of information [11].

The omission of studies with incomplete information may affect the results of an NMA through a loss of statistical power and bias in the final estimates [11]. By far the most common approach to the missing data is to omit those studies with the missing data from the analysis. This approach is known as the complete-case analysis or list-wise deletion. Some authors suggest that, implicitly, in this case, all missing data are considered as missing completely at random (MCAR) [12]. If the missing data are MCAR, they are entirely independent of observed and unobserved data [13]. This assumption is not verifiable for data about candidate studies to be included in an NMA model; the missing information may be considered a function of the missing data itself and may be influenced by other variables related to study characteristics, for example, sample size [14] or the effect size reported in the study [15]. In fact, significant results are more likely to be reported entirely in clinical trial publications [10, 16].

This study investigates the problem of missing data imputation in an NMA on continuous outcomes where the treatment effect is expressed as delta (i.e., mean variation) between baseline and follow-up for each treatment included in the network. In this framework, different imputation strategies may be performed depending on the missing information. Some studies may report an average effect related to baseline and follow-up assessment and omit the mean treatment effect (with corresponding variability measure) in terms of delta variation. In this case,

different strategies are provided by the literature, mostly based on the imputation of the pre-post study correlation [17] that is used to obtain the delta variability measure [18]. The correlation may be imputed considering other comparable studies [19] or using a conservative estimate of 0.5 [19]. Among Bayesian alternatives, Abrams and colleagues suggest the exploitation of complete available evidence composed by baseline, follow-up and delta data to obtain a posterior distribution for the correlation [17].

If a study just reports the delta without related variability and mean effect at baseline and follow-up, the imputation strategy is reduced to the imputation of standard deviation (SD). Several approaches have been proposed, among them maximum SD imputation [20], arithmetic mean [21], linear regression [22], the coefficient of variation [23], bootstrap method [24, 25] and multiple imputations [13, 26].

The purpose of this study is to verify the robustness of a Bayesian NMA model concerning some common imputation strategies on data characterized by different levels of between-trial heterogeneity. A simulation study has been performed to assess the performance of the method in ranking treatments accurately when the network includes imputed data. Data have been simulated referring to the effects of treatments, as reported in the literature, against knee osteoarthritis.

MATERIALS AND METHODS

Imputation methods

Let x_{ij} and y_{ij} denote the i_{th} outcome in the j_{th} group of the study at baseline and follow-up, respectively. Let d_{ij} the mean difference in change from baseline to follow up i_{th} in the group of the j_{th} study, and $\sigma^2_{dij}(\rho)$ its associated variance. The following approaches are evaluated in this study:

Maximum SD imputation

This method is based on a direct SD imputation from studies included in the meta-analysis [20]. Although other more sophisticated methods are available [27], this approach is commonly used in practice to obtain conservative estimates [8].

Correlation approach

The variability of mean change score d_{ij} defined $V(d_{ij})$ can be calculated as:

$$V(d_{ij}) = V(x_{ij}) + V(y_{ij}) - 2\rho \sqrt{V(x_{ij})V(y_{ij})}$$

In this setting, using complete evidence composed by

baseline, follow-up and delta data, a meta-analysis of the Fisher transformations, S_{ij} is considered. Therefore, data derived from J studies having complete data are used to derive the posterior distribution for correlation.

The Fisher transformation is [17]

$$S_{ij} \sim N[\eta_{ij}, v_{ij}]$$

The mean of the transformation is a normally distributed random variable with a normal distribution on the mean and a Gamma distribution on the variance:

$$\eta_{ij} \sim N[\delta, \omega^2] \quad for \quad j = 1, \dots, J$$

$$\delta \sim N[0, 100^2] \ and \ \omega^2 \sim Gamma[0.001, 0.001]$$

The correlation is strictly related to the Fisher transformation via the δ parameter, as shown below:

$$\rho = \frac{e^{2\delta} - 1}{e^{2\delta} + 1}$$

The use of a fully Bayesian approach for the imputation of correlation is suitable not only because it leads the researcher to include in the computation pertinent study information but also because it takes into account a source of uncertainty in parameter estimation [17].

All these factors eliminate the need for sensitivity analyses with respect to an arbitrary choice of the within-subject correlations [17].

Simulation framework

Seven hundred full databases, including study-level baseline (x_{ij}), follow-up (y_{ij}), and delta variation (d_{ij}), including 50 two-arm trials of size n_{ij}, are simulated (*j*=2 pairwise comparisons). Trial data are simulated considering as an outcome the delta variation in the Western Ontario and McMaster Universities index (WOMAC) [28], widely used to assess the severity of symptoms in knee osteoarthritis. Baseline data are obtained by sampling from bounded 0-100 normal distributions with mean 41.8 and SD 21.5, to mimic the support of WOMAC score [29]. Delta variation data are simulated from normal distributions with parameters provided by a literature review about six non-steroidal anti-inflammatory drugs (NSAIDs) against symptoms of knee osteoarthritis, as shown in Table 1.

For each simulated trial, treatments involved have been randomly drawn from the list of treatments shown in Table 1. Follow-up variability data (standard error of the mean) are derived from the generated delta and baseline variability measures, using the inverse formula, setting

hypotheses on pre-post correlation (ρ) and considering, in each scenario, a sequence of ρ values from 0.3 to 0.95 by in increments of 0.05. The sample size n_j is obtained by sampling from a uniform distribution bounded 50-100, reflecting the general framework of the sample size in rheumatological trials [30].

Between-trial heterogeneity τ has been included as a variability measure by following, for each simulation setting, a wide range of heterogeneity values consisting in a sequence from 0.1 to 5 by 0.1. Each simulation scenario provides different combinations of between-trial heterogeneity and pre-post correlation, totalling 700 scenarios.

More details about the simulation plan are shown in Figure 1.

Imputed data generation process

Imputed database I. In the first case, information about delta variation is randomly removed from the full database, leaving only baseline and follow-up data. Out of 50 studies, the variability information has been randomly removed in 10 studies. Variability of mean change has been imputed using the correlation method.

Imputed database II. In this case, information at baseline and follow-up are also removed, and then delta variability is imputed with maximum SD method.

Bayesian NMA analysis

A multiple-treatment Bayesian NMA, with random effects and a uniform (0,5) prior distribution on the heterogeneity parameter [31], has been performed on each simulated dataset consisting of 50 studies, using an arm-based approach [19]. An arm-based approach models the arm-level data referring to a specific treatment effect within each trial. For comparison, a contrast-based approach would be based on the parametrization of the contrast trial level summary (for example, the mean difference in treatment effect within the trial). The results provided by both approaches are very similar [32]. An arm-level approach, compared to contrast-based ones, leads to the incorporation of multi-arm trials without the necessity to employ multivariate distribution to determine specific treatment effects [32].

An MCMC approach was used to sample from the posterior distribution estimated by the NMA model. The sampling processes involved 200,000 iterations, including a burn in of 20,000 steps, from 4 chains. The results are compared in term of ranking probability [33].

For each scenario, the bias and the standard deviation of the first rank probability estimates, between full and imputed databases, have been computed to investigate the robustness of NMA results with respect to the considered imputation methods.

FIGURE 1. Simulation plan: the treatments are sampled from the drugs reported in Table 1, the sample size, instead are drawn from a uniform 0-100 distribution.
The delta data are sampled from information provided in Table 1, and the baseline data are drawn from a 0-100 Normal distribution having mean 41.8 and SD 21.5.
The imputed databases (I) have been obtained randomly removing, from the full database, information about delta data.
The Imputed databases (II) have been obtained randomly removing baseline and follow up information.
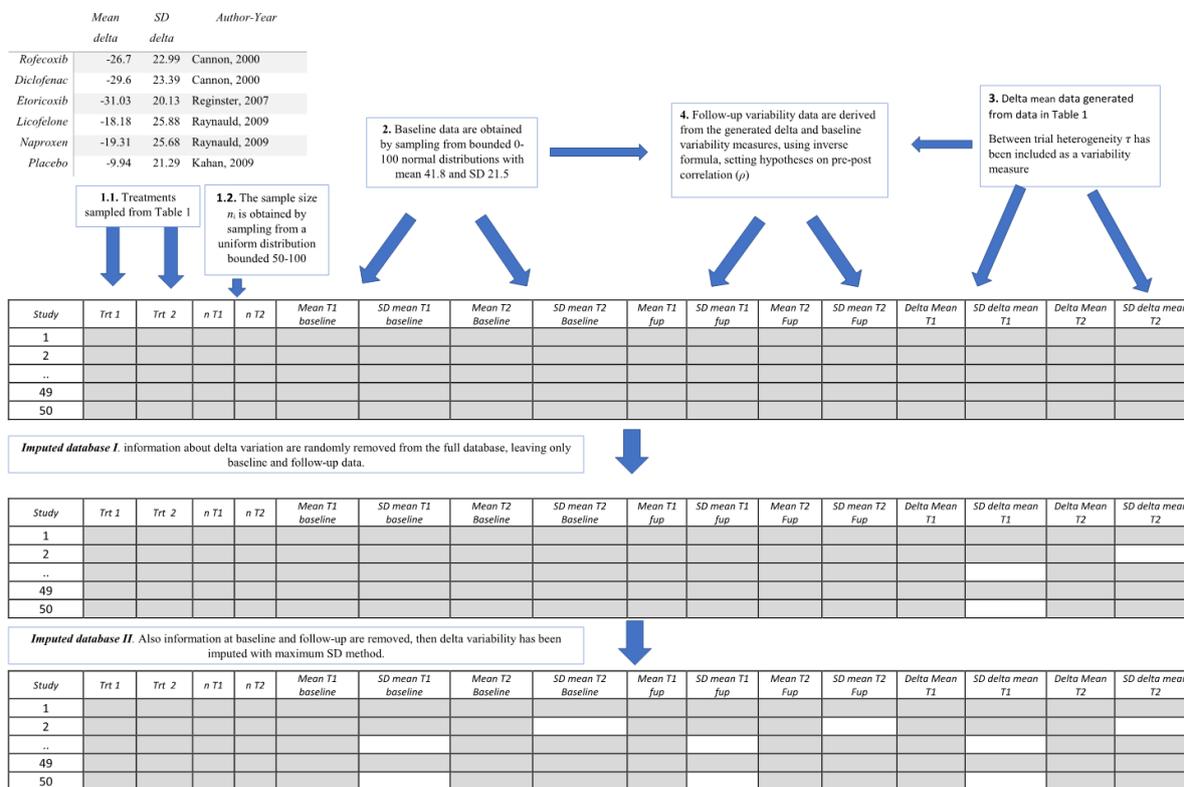
TABLE 1. Delta Mean and Standard deviation of NSAIDs treatment derived by the literature. The parameters are used to generate data for delta variation treatment effect.

|	Mean delta	SD delta	Author-Year
Rofecoxib	-26.7	22.99	Cannon, 2000 [29]
Diclofenac	-29.6	23.39	Cannon, 2000 [29]
Etoricoxib	-31.03	20.13	Reginster, 2007 [45]
Licofelone	-18.18	25.88	Raynauld, 2009 [46]
Naproxen	-19.31	25.68	Raynauld, 2009 [46]
Placebo	-9.94	21.29	Kahan, 2009 [47]

RESULTS

The network plot (Figure 2) is a graphical representation of the overall evidence leading to a concise description of its characteristics. Each node of the graph represents the interventions, and edges represent the available direct evidence between treatments. The size of the node is proportional to the sample size in each treatment, and line thickness is proportional to the number of pairwise comparisons between treatments. The network structure is the same in every scenario. Observing the structure of the network, it is possible to see that no indirect evidence is included in the meta-analysis. In fact, all comparisons report direct evidence in the network structure shown in Figure 2.

Results are synthesized in term of first rank probabilities for each treatment reported in Table 2. The simulation results are coherent for each scenario for all the imputation methods. The first rank probabilities are higher for etoricoxib followed by diclofenac and rofecoxib and they are always equal to zero for licofelone, naproxen, and placebo. The ranking provided by the models is robust

FIGURE 2. Network plot of the simulated evidence. Each node represents a treatment included in the network. The line width is proportional to the number of studies reporting the specific pairwise comparison.
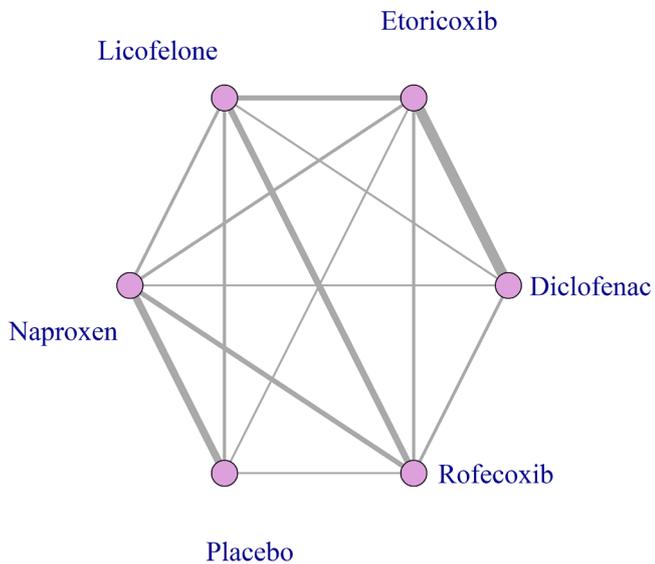

with respect to different imputation methods in several heterogeneity/correlation settings.

The robustness of the final inference to missing data imputation has been evaluated considering bias and standard deviation within scenario (Table 3). The results are similar across different correlation settings. Smallest values of bias and SD are obtained from smaller values of heterogeneity. The bias increases as heterogeneity increases. This pattern is evident only in the presence of the lowest and highest values of heterogeneity, as shown in Figure 3. Additionally, it may be observed that the lowest bias and SDs are obtained, not only in the case of the lowest heterogeneity setting but also for heterogeneity values of 2.5 to 3.5, near the expectation of the uniform prior distribution chosen on heterogeneity parameters (Figure 3).

DISCUSSION

Imputation methods are generally viewed as the preferred analytical approach to the missing data problem in NMA [11, 34]. When no particular imputation method is preferable to others, a sensitivity analysis of NMA results concerning different imputation strategies is advocated [35]. It should be noted that few NMA studies report the

FIGURE 3. Bias and Standard deviation within scenarios for different correlation settings. In Bayesian NMA prior to heterogeneity τ, has been chosen as a Uniform (0,5) distribution.
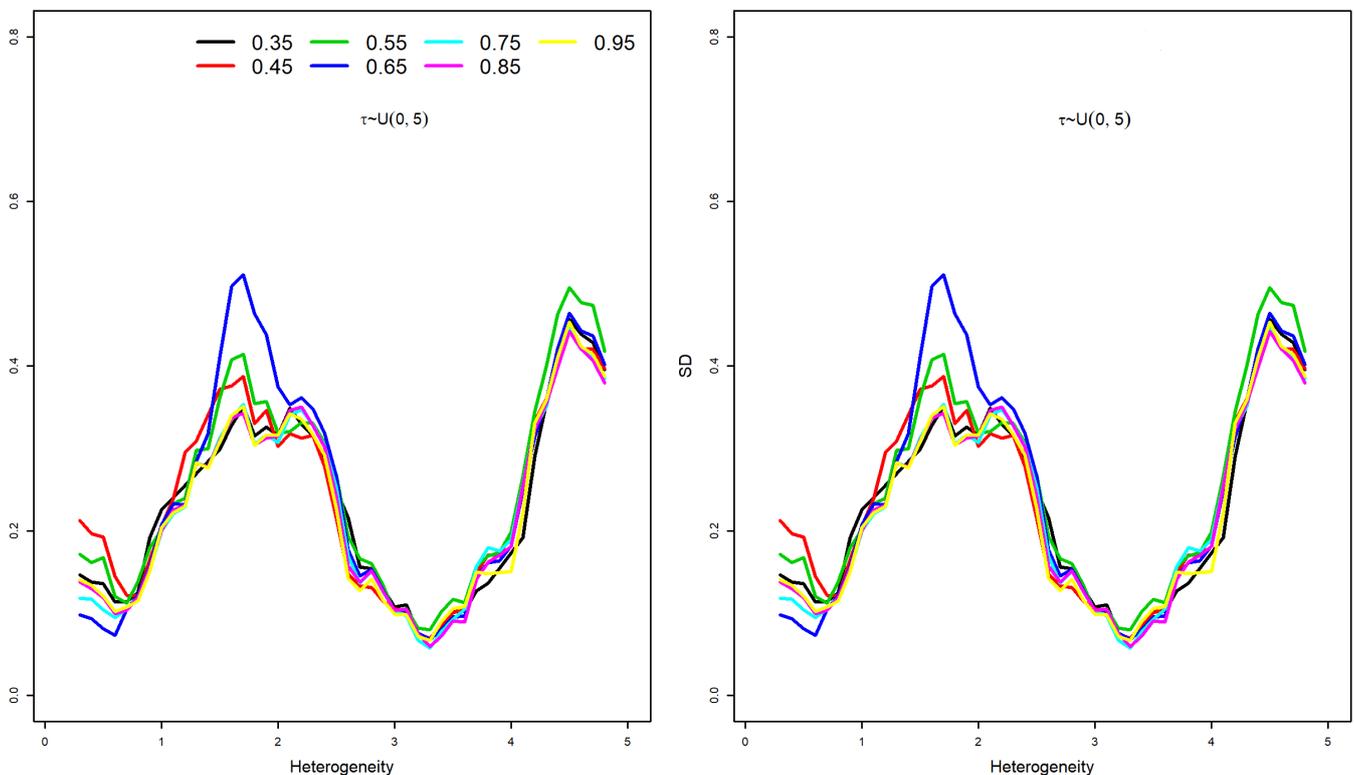
TABLE 2. First Rank probabilities for each treatment in each scenario.

Diclofenac	Etoricoxib	Licofelone	Naproxen	Placebo	Rofecoxib	τ	ρ	Imputation
0.191	0.792	0	0	0	0.017	0.1	0.3	Full DB
0.197	0.786	0	0	0	0.017	0.1	0.3	Max SD
0.194	0.790	0	0	0	0.017	0.1	0.3	Abrams
0.191	0.792	0	0	0	0.017	0.1	0.5	Full DB
0.197	0.786	0	0	0	0.017	0.1	0.5	Max SD
0.194	0.788	0	0	0	0.018	0.1	0.5	Abrams
0.191	0.792	0	0	0	0.017	0.1	0.7	Full DB
0.197	0.786	0	0	0	0.017	0.1	0.7	Max SD
0.192	0.789	0	0	0	0.019	0.1	0.7	Abrams
0.191	0.792	0	0	0	0.017	0.1	0.9	Full DB
0.197	0.786	0	0	0	0.017	0.1	0.9	Max SD
0.195	0.787	0	0	0	0.018	0.1	0.9	Abrams
0.161	0.828	0	0	0	0.011	1	0.3	Full DB
0.169	0.818	0	0	0	0.013	1	0.3	Max SD
0.169	0.819	0	0	0	0.012	1	0.3	Abrams
0.161	0.828	0	0	0	0.011	1	0.5	Full DB
0.169	0.818	0	0	0	0.013	1	0.5	Max SD
0.166	0.822	0	0	0	0.012	1	0.5	Abrams
0.161	0.828	0	0	0	0.011	1	0.7	Full DB
0.169	0.818	0	0	0	0.013	1	0.7	Max SD
0.163	0.825	0	0	0	0.012	1	0.7	Abrams
0.161	0.828	0	0	0	0.011	1	0.9	Full DB
0.169	0.818	0	0	0	0.013	1	0.9	Max SD
0.166	0.822	0	0	0	0.012	1	0.9	Abrams
0.232	0.729	0	0	0	0.039	2	0.3	Full DB
0.236	0.730	0	0	0	0.033	2	0.3	Max SD
0.233	0.732	0	0	0	0.035	2	0.3	Abrams
0.232	0.729	0	0	0	0.039	2	0.5	Full DB
0.236	0.730	0	0	0	0.033	2	0.5	Max SD
0.230	0.734	0	0	0	0.036	2	0.5	Abrams
0.232	0.729	0	0	0	0.039	2	0.7	Full DB
0.236	0.730	0	0	0	0.033	2	0.7	Max SD
0.230	0.733	0	0	0	0.037	2	0.7	Abrams
0.232	0.729	0	0	0	0.039	2	0.9	Full DB
0.236	0.730	0	0	0	0.033	2	0.9	Max SD
0.229	0.735	0	0	0	0.036	2	0.9	Abrams
0.091	0.892	0	0	0	0.018	3	0.3	Full DB
0.082	0.902	0	0	0	0.016	3	0.3	Max SD
0.085	0.898	0	0	0	0.017	3	0.3	Abrams
0.091	0.892	0	0	0	0.018	3	0.5	Full DB

A simulation scenario consists of different combination of heterogeneity () and correlation (ρ = 0.3, 0.5, 0.7, 0.9).
First Rank probability estimates are reported for an NMA computed on Full Databases and imputed databases (Maximum SD and Abrams method).

TABLE 2 (CONTINUED). First Rank probabilities for each treatment in each scenario.

Diclofenac	Etoricoxib	Licofelone	Naproxen	Placebo	Rofecoxib	τ	ρ	Imputation
0.082	0.902	0	0	0	0.016	3	0.3	Max SD
0.085	0.898	0	0	0	0.017	3	0.3	Abrams
0.091	0.892	0	0	0	0.018	3	0.5	Full DB
0.082	0.902	0	0	0	0.016	3	0.5	Max SD
0.086	0.897	0	0	0	0.018	3	0.5	Abrams
0.091	0.892	0	0	0	0.018	3	0.7	Full DB
0.082	0.902	0	0	0	0.016	3	0.7	Max SD
0.087	0.897	0	0	0	0.016	3	0.7	Abrams
0.091	0.892	0	0	0	0.018	3	0.9	Full DB
0.082	0.902	0	0	0	0.016	3	0.9	Max SD
0.086	0.898	0	0	0	0.017	3	0.9	Abrams
0.112	0.842	0	0	0	0.046	4	0.3	Full DB
0.122	0.829	0	0	0	0.049	4	0.3	Max SD
0.118	0.833	0	0	0	0.048	4	0.3	Abrams
0.112	0.842	0	0	0	0.046	4	0.5	Full DB
0.122	0.829	0	0	0	0.049	4	0.5	Max SD
0.113	0.839	0	0	0	0.047	4	0.5	Abrams
0.112	0.842	0	0	0	0.046	4	0.7	Full DB
0.122	0.829	0	0	0	0.049	4	0.7	Max SD
0.113	0.838	0	0	0	0.050	4	0.7	Abrams
0.112	0.842	0	0	0	0.046	4	0.9	Full DB
0.122	0.829	0	0	0	0.049	4	0.9	Max SD
0.112	0.840	0	0	0	0.048	4	0.9	Abrams
0.181	0.777	0	0	0	0.042	5	0.3	Full DB
0.191	0.763	0	0	0	0.045	5	0.3	Max SD
0.182	0.774	0	0	0	0.044	5	0.3	Abrams
0.181	0.777	0	0	0	0.042	5	0.5	Full DB
0.191	0.763	0	0	0	0.045	5	0.5	Max SD
0.184	0.771	0	0	0	0.045	5	0.5	Abrams
0.181	0.777	0	0	0	0.042	5	0.7	Full DB
0.191	0.763	0	0	0	0.045	5	0.7	Max SD
0.183	0.771	0	0	0	0.046	5	0.7	Abrams
0.181	0.777	0	0	0	0.042	5	0.9	Full DB
0.191	0.763	0	0	0	0.045	5	0.9	Max SD
0.186	0.769	0	0	0	0.045	5	0.9	Abrams

A simulation scenario consists of different combination of heterogeneity () and correlation (ρ = 0.3, 0.5, 0.7, 0.9).

First Rank probability estimates are reported for an NMA computed on Full Databases and imputed databases (Maximum SD and Abrams method).

results of sensitivity analysis after imputation [21, 36, 37], especially for variance imputation on continuous outcomes.

Some authors have investigated the performance of the missing variance imputation strategies on continuous data considering a pairwise meta-analysis setting [38], showing the substantial robustness of results in several scenarios, especially when a random-effect meta-analysis has been performed. This study, instead, verifies the robustness of a Bayesian NMA model concerning different imputation strategies through simulations and reaches similar conclusions which are known to hold for pairwise meta-analysis in other research settings [39]. Our findings

TABLE 3. Mean of Bias (Absolute Values) and Standard Deviation (*100) between treatments in each simulation scenario.

Bias	SD	τ	ρ
0.137	0.125	0.1	0.3
0.163	0.136	0.1	0.5
0.146	0.152	0.1	0.7
0.188	0.136	0.1	0.9
0.319	0.195	1	0.3
0.276	0.17	1	0.5
0.221	0.172	1	0.7
0.278	0.171	1	0.9
0.056	0.172	2	0.3
0.103	0.206	2	0.5
0.084	0.21	2	0.7
0.108	0.214	2	0.9
0.279	0.177	3	0.3
0.26	0.178	3	0.5
0.26	0.177	3	0.7
0.277	0.176	3	0.9
0.348	0.213	4	0.3
0.25	0.222	4	0.5
0.276	0.247	4	0.7
0.232	0.233	4	0.9
0.286	0.241	5	0.3
0.33	0.235	5	0.5
0.33	0.248	5	0.7
0.373	0.237	5	0.9

The Bias has been computed as the mean of the differences between first rank probability estimate on the imputed database and the same estimate obtained on the full database.
The Standard deviation has been computed on the first rank estimates computed by NMA in each scenario.

show that the first rank probability bias is small over a broad range of imputation scenarios, showing the substantial robustness of final inference.

The results of our study are in line with previous findings that suggest, in the presence of missing variance data, we should consider an imputation method rather than exclude the study from the analysis [40]. In fact, we have to consider that it is difficult to suppose that the missingness in variance study-level data may be caused by MAR mechanism [14, 15].

Another noteworthy aspect of this paper is the sensitivity of NMA after imputation in different heterogeneity scenarios across studies. Heterogeneity is a crucial element in a meta-analysis. We have to take into account that heterogeneity is a source of variability, related to differences across the trial populations and considered interventions, which can occur in many studies that are candidates for inclusion in a meta-analysis [41]. In addition to missing data, heterogeneity between trials is a component that may affect final inference. Heterogeneity between trials indicates a source of variability in treatment effects that, if misspecified (i.e. poorly identified by the prior distribution) in an NMA model, results in a limitation in the external validity of the results [33], downing the possibilities of generalizing inferential results to other situations.

This additional source of variability may occur when studies differing in trial characteristics, outcome definition, inclusion criteria or follow-up duration are included in the same meta-analysis [42]. Opportune statistical methods are available to handle this additional source of variability, for example, a random effect NMA may be implemented in a Frequentist or a Bayesian setting [43]. In practical approaches, the Bayesian methods for mixed-treatment-comparison meta-analyses are widely used compared to frequentist NMA [43].

The literature on the relationship between continuous missing data imputation methods and heterogeneity focuses on pairwise meta-analysis [34]. Specifically, scholars suggest that random-effect meta-analysis, even in cases of missing data, might be replaced with fixed-effect meta-regression to handle a source of heterogeneity that is not influenced by the known source of variability across studies. Instead, in the setting of a Bayesian NMA, different authors suggest some good practices to deal with heterogeneity, especially to define priors on heterogeneity across studies [31, 44]. In this work, in line with the literature, we confirm that this source of variability is an important issue to consider. In fact, the correct specification of the prior defined on the heterogeneity leads to more robust results to imputation, especially if the prior on heterogeneity value has an expectation near the true value of the variability across studies.

Study limitations

In further research, it may be useful to assess the performance of the NMA with respect to missing imputation, considering a smaller number of trials in the network, and including a different proportion of direct and indirect comparison.

CONCLUSIONS

Inference provided by NMA models is robust with respect to different imputation strategies or considering different combinations of heterogeneity and correlation when the proportion of missing data is 20%. The Bayesian NMA leads to a better performance also when data are imputed, especially when the extent of heterogeneity across studies is limited. The definition of the prior on

heterogeneity is an important issue in this context: the inference performed with an NMA on imputed data is more robust to missing value imputation if the prior distribution is well specified.

Conflict of interest

None to declare.

References

1. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21(16):2313–24.

2. Salanti G, Ades A, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011;64(2):163–71.

3. Tonin FS, Rotta I, Mendes AM, Pontarolo R. Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. Pharm Pract 2017;15(1).

4. Caldwell DM, Dias S, Welton NJ. Extending treatment networks in health technology assessment: how far should we go?. Value Health 2015;18(5):673–81.

5. Bafeta A, Trinquart L, Seror R, Ravaud P. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. BMJ 2013;347:f3675.

6. Bafeta A, Trinquart L, Seror R, Ravaud P. Reporting of results from network meta-analyses: methodological systematic review. BMJ 2014;348:g1741.

7. Jones B, Roger J, Lane PW, et al. Statistical approaches for conducting network meta-analysis in drug development. Pharm Stat 2011;10(6):523–31.

8. Fu R, Vandermeer BW, Shamliyan TA, et al. Handling continuous outcomes in quantitative synthesis. 2013.

9. Dias S, Welton NJ, Sutton AJ, Ades A. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. 2011.

10. 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):2457–65.

11. Batson S, Burton H. A systematic review of methods for handling missing variance data in meta-analyses of interventions in type 2 diabetes mellitus. PloS One 2016;11(10):e0164827.

12. White IR, Higgins J, Wood AM. Allowing for uncertainty due to missing data in meta-analysis—Part 1: Two-stage methods. Stat Med 2008;27(5):711–27.

13. Little RJ, Rubin DB. Single imputation methods. Stat Anal Missing Data Second Ed 2002;59–74.

14. Allison PD. Missing data: Quantitative applications in the social sciences. Br J Math Stat Psychol 2002;55(1):193–6.

15. Wiebe N, Vandermeer B, Platt RW, Klassen TP, Moher D, Barrowman NJ. A systematic review identifies a lack of standardization in methods for handling missing variance data. J Clin Epidemiol 2006;59(4):342–53.

16. Chan A-W, Krleža-Jeri- K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. Can Med Assoc J 2004;171(7):735–40.

17. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. Stat Med 2005;24(24):3823–44.

18. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. Vol. 4. John Wiley & Sons; 2011.

19. Dias S, Welton N, Sutton A, A A NICE D. technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. ScHARR Univ Sheff Decis Support Unit. 2011;

20. Mark W.. Lipsey, Wilson DB. Practical meta-analysis. Vol. 49. Sage publications Thousand Oaks, CA; 2001.

21. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. J Clin Epidemiol 1992;45(7):769–73.

22. Buck SF. A method of estimation of missing values in multivariate data suitable for use with an electronic computer. J R Stat Soc Ser B Methodol 1960;302–6.

23. Speidel B. Effective Care of the Newborn Infant. Arch Dis Child 1992;67(11):1415.

24. Zhu W. Making bootstrap statistical inferences: a tutorial. Res Q Exerc Sport 1997;68(1):44–55.

25. Kelley GA, Kelley KS, Tran ZV. Exercise and bone mineral density in men: a meta-analysis. J Appl Physiol 2000;88(5):1730–6.

26. Kasiske BL, Lakatua J, Ma JZ, Louis TA. A meta-analysis of the effects of dietary protein restriction on the rate of decline in renal function. Am J Kidney Dis 1998;31(6):954–61.

27. Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P, et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. BMC Med Res Methodol 2018;18(1):25.

28. Bellamy N. WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. J Rheumatol 2002;29(12):2473–6.

29. Cannon GW, Caldwell JR, Holt P, et al. Rofecoxib, a specific inhibitor of cyclooxygenase 2, with clinical efficacy comparable with that of diclofenac sodium: Results of a one-year, randomized, clinical trial in patients with osteoarthritis of the knee and hip. Arthritis Rheumatol 2000;43(5):978–87.

30. Celik S, Yazici Y, Yazici H. Are sample sizes of randomized clinical trials in rheumatoid arthritis too large? Eur J Clin Invest 2014;44(11):1034–44.

31. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat Med 2015;34(6):984–98.

32. Hawkins N, Scott DA, Woods B. 'Arm-based'parameterization for network meta-analysis. Res Synth Methods 2016;7(3):306–13.

33. Dias S, Sutton AJ, Welton NJ, Ades A. Evidence synthesis for decision making 3: heterogeneity—subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making 2013;33(5):618–40.

34. Chowdhry AK, Dworkin RH, McDermott MP. Meta-analysis with missing study-level sample variance data. Stat Med 2016;35(17):3021–32.

35. Guevara JP, Wolf FM, Grum CM, Clark NM. Effects of educational interventions for self management of asthma in children and adolescents: systematic review and meta-analysis. BMJ 2003;326(7402):1308–9.

36. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol 2006;59(1):7–10.

37. Philbrook HT, Barrowman N, Garg A. Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. J Clin Epidemiol 2007;60(3):228–40.

38. Idris N, Abdullah M, Tolos S. Performance of selected imputation techniques for missing variances in meta-analysis. In IOP Publishing 2013. p. 012037.

39. Greco T, Biondi-Zoccai G, Gemma M, Guérin C, Zangrillo A, Landoni G. How to impute study-specific standard deviations in meta-analyses of skewed continuous endpoints?. World J Meta-Anal 2015;3(5):215–24.

40. Batson S, Burton H. A systematic review of methods for handling missing variance data in meta-analyses of interventions in type 2 diabetes mellitus. PloS One 2016;11(10):e0164827.

41. Schmitz S, Kandala N, Senn S.Aan exploration of the heterogeneity in a network meta-analysis explained by dosing differences. Value Health 2016;19(3):A89.

42. Greco T, Biondi-Zoccai G, Saleh O, Pasin L, Cabrini L, Zangrillo A, et al. The attractiveness of network meta-analysis: a comprehensive systematic and narrative review. Heart Lung Vessels 2015;7(2):133.

43. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011;14(4):429–37.

44. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med 2005;24(15):2401–28.

45. Reginster J-Y, Malmstrom K, Mehta A, et al. Evaluation of the efficacy and safety of etoricoxib compared with naproxen in two, 138-week randomised studies of patients with osteoarthritis. Ann Rheum Dis 2007;66(7):945–51.

46. Raynauld J-P, Martel-Pelletier J, Bias P, et al. Protective effects of licofelone, a 5-lipoxygenase and cyclo-oxygenase inhibitor, versus naproxen on cartilage loss in knee osteoarthritis: a first multicentre clinical trial using quantitative MRI. Ann Rheum Dis 2009;68(6):938–47.

47. Kahan A, Uebelhart D, De Vathaire F, Delmas PD, Reginster J. Long-term effects of chondroitins 4 and 6 sulfate on knee osteoarthritis: The study on osteoarthritis progression prevention, a two-year, randomized, double-blind, placebo-controlled trial. Arthritis Rheumatol 2009;60(2):524–33.