Allowing for uncertainty due to missing continuous outcome data in pairwise and network meta-analysis

Dimitris Mavridis, Ian R. White, Julian P. T. Higgins, Andrea Cipriani and Georgia Salanti

Missing outcome data are commonly encountered in randomized controlled trials and hence may need to be addressed in a meta-analysis of multiple trials. A common and simple approach to deal with missing data is to restrict analysis to individuals for whom the outcome was obtained (complete case analysis). However, estimated treatment effects from complete case analyses are potentially biased if informative missing data are ignored. We develop methods for estimating meta-analytic summary treatment effects for continuous outcomes in the presence of missing data for some of the individuals within the trials. We build on a method previously developed for binary outcomes, which quantifies the degree of departure from a missing at random assumption via the informative missingness odds ratio. Our new model quantifies the degree of departure from missing at random using either an informative missingness difference of means or an informative missingness ratio of means, both of which relate the mean value of the missing outcome data to that of the observed data. We propose estimating the treatment effects, adjusted for informative missingness, and their standard errors by a Taylor series approximation and by a Monte Carlo method. We apply the methodology to examples of both pairwise and network meta-analysis with multi-arm trials. © 2014 The Authors. Statistics in Medicine Published by John Wiley & Sons Ltd.

Keywords: informative missing; mixed treatment comparison; sensitivity analysis

1. Introduction

Even in carefully designed and conducted randomized controlled trials (RCTs), patients may be lost to follow up. Exclusion of missing participants from the analysis may, under some circumstances, reduce precision in the estimated treatment effects and produce biased results. In mental health trials, for instance, the dropout rate is generally high and may exceed 50% [1]. It has been shown that patients in a placebo arm may be more likely to drop out because of lack of efficacy, and patients in an active treatment arm may be more likely to drop out because of severe adverse events [2]. This corresponds to informatively missing (IM) data [3]: reasons for dropout are associated with the outcome measured, given all observed data. Bias is especially likely when reasons for dropout are also associated with the treatment assigned. Failure to account for the missing outcome data in the evidence provided from RCTs and their meta-analysis can lead to biased conclusions.

The intention-to-treat principle is widely accepted as the most appropriate way to analyze RCT data, as it avoids biases caused by participants switching interventions or dropping out of the trial. The intention-to-treat principle can only be followed fully if there are complete outcome data for all randomized participants.
participants. In an RCT, this is possible when appropriate analysis strategies are employed for the missing outcome data [4].

When conducting a meta-analysis, however, the data available are usually much simpler: for each arm, just the number of randomized individuals, the number with observed outcome, and simple summaries of the observed outcomes. In this setting, it is common practice to ignore the missing data or consider that the missing data have been handled at the study level (e.g., via an imputation approach). In this paper, however, we assume there has been no imputation at the study level. Ignoring missing outcome data in a complete case (CC) analysis is valid in general when data are missing at random (MAR). For this data structure, MAR means that the probability of data being missing does not depend on the outcome within randomized groups, which is equivalent to missing completely at random within randomized groups. Meta-analysts typically do not have access to the reasons for missing outcome data, and the MAR assumption cannot be explored empirically. Consequently, a sensitivity analysis is the only viable way to evaluate the effect of different scenarios for the missing data mechanism [5]. A review of methods currently used for handling missing outcome data in meta-analysis can be found in [6].

Typical sensitivity analyses include the best-case and worst-case scenarios, defining the extremes that could have occurred had all outcomes been observed. These two extremes may be unrealistic in practice. Gamble and Hollis suggested that studies for which there is a big discrepancy between these two extremes should be downweighted [7]. A more elegant way to undertake sensitivity analysis is via a pattern-mixture model that makes hypotheses for the unobserved data given the observed data [8]. For the case of missing binary outcomes, White et al. presented a pattern-mixture model where the degree of departure from the MAR assumption is quantified by the informative missingness odds ratio (IMOR) [9]. The IMOR is defined as the ratio of the odds of success in the missing data to the corresponding odds in the observed data. Higgins et al. provide an overview of imputation strategies for binary data, including the IMOR approach and impute risks among missing participants for each group based on available reasons for missingness [10]. An advantage of the IMOR model is that it is plausibly unaffected by the amount of missing data, unlike the similar measure based on the risk ratio suggested by Magder [11]. Spineli et al. extended the IMOR to a network meta-analysis (NMA) setting and explored the impact of missing outcome data on inferences about the relative effectiveness of several competing treatments [12]. We see two major advantages of the IMOR approach. First, it does not aim to estimate the missing values but, in the spirit of Shafer and Graham [13], aims to make valid inference on the parameters of interest. Second, it accounts for the uncertainty induced by missing outcome data, unlike naïve approaches that consider the imputed values as if they were fully observed [14].

Little work has been carried out to date for quantitative (continuous) missing outcome data in meta-analysis. Standard approaches include CC analysis and imputation of the observed means for the missing participants [15]. This paper proposes a new pattern-mixture model for meta-analysis of trials with a single continuous outcome measured at one time point. Our model enables estimation of summary effects while accounting for uncertainty in the outcome of the participants who dropped out. We allow the treatment effect to be measured as a mean difference (MD), as a standardized MD (SMD), or as a ratio of means. Instead of the IMOR, we quantify the degree of departure from MAR by the informative missingness difference of means (IMDoM) or by the informative missingness ratio of means (IMRoM). We address both pairwise and NMA, and we extend the methods to handle multi-arm trials.

The paper is laid out as follows. In Section 2, we present the pattern-mixture model, introduce the measures of departure from MAR, and propose two estimation methods. In Section 3, we present the pairwise and NMA models. In Section 4, the proposed methodology is illustrated in two data sets for pairwise and NMA. We conclude with a discussion in Section 5.

2. Methods

2.1. Notation and model definition

We suppose there are $m_{ij} + n_{ij}$ participants randomized in arm of study $i$, with $n_{ij}$ participants providing outcome data and $m_{ij}$ participants for whom outcome data are not available. For individual $k = 1, \ldots, m_{ij} + n_{ij}$ in arm $j$, let $Y_{ik}$ be the (possibly missing) outcome, and let the indicator variable $R_{ijk}$ assume value 1 if the outcome has been observed and 0 otherwise. In this section, we assume two-arm trials, so $j = C$ for control or $T$ for treatment; In Section 3.2, we extend the model to multi-arm trials.
We write our model as
\[ P(R_{ijk} = 1) = \pi_{ij}, \]
\[ E(Y_{ijk} | R_{ijk} = 1) = \chi_{ij}^{obs}, \]
\[ E(Y_{ijk} | R_{ijk} = 0) = \chi_{ij}^{miss}. \]

We use Greek letters \( \pi, \chi \) to denote parameters and Latin letters \( p, x \) for their sample estimates. We can estimate \( \pi_{ij} \) and \( \chi_{ij}^{obs} \) directly from the data, whereas \( \chi_{ij}^{miss} \) cannot be estimated from the data. The probability that an individual provides outcome data, \( \pi_{ij} \), can be estimated by \( p_{ij} = \frac{n_{ij}}{n_i + n_j} \), and \( \chi_{ij}^{obs} \) is estimated by the sample mean \( \chi_{ij}^{obs} \) with variance \( \frac{s_{ij}^2}{n_{ij}} \), where \( s_{ij} \) is the standard deviation of the outcome among observed participants.

The unconditional expectation of the outcome in each study arm is
\[ E(Y_{ijk}) \equiv \chi_{ij}^{tot} = \pi_{ij}\chi_{ij}^{obs} + (1 - \pi_{ij})\chi_{ij}^{miss}. \]

The unconditional means \( \chi_{ij}^{tot} \) are contrasted to obtain the relative treatment effect in each study, which is defined as the difference
\[ \beta_i \equiv f(\chi_{iT}) - f(\chi_{iC}), \]
where \( f \) is a link function that determines the effect measure. \( \beta_i \) is the MD if \( f(u) = u \); the logarithm of the ratio of means (ln RoM) if \( f(u) = \ln(u) \) \cite{16}; and the SMD if \( f(u_i) = \frac{u_i}{s_i} \), where \( s_i = \sqrt{\frac{(n_{ij} - 1)s_{ij}^2 + (n_i + n_j - 1)s_i^2}{n_i + n_j + n_{ij} - 2}} \) is the usual pooled standard deviation.

### 2.2. Quantifying informative missingness: connecting mean outcome values between observed and missing data

We quantify informative missingness (IM) by making assumptions about the relationship between the outcome means from the observed data and missing data. Specifically, we set
\[ \lambda_{ij} = g(\chi_{ij}^{miss}) - g(\chi_{ij}^{obs}) \]
so that the mean value in the missing data in study \( i \) and arm \( j \) is related to the corresponding mean of the observed data by an unknown parameter \( \lambda_{ij} \). Equation 1 implies that \( \chi_{ij}^{miss} = g^{-1}(\lambda_{ij} + g(\chi_{ij}^{obs})) \). We consider two options for the link function \( g \):

1. If \( g \) is the identity function, \( \lambda_{ij} = \chi_{ij}^{miss} - \chi_{ij}^{obs} \) is the IMDoM.
2. If \( g \) is the ln function, \( \exp(\lambda_{ij}) = \chi_{ij}^{miss} / \chi_{ij}^{obs} \) is the IMRoM.

From Equations 1 and 3, it follows that
\[ \chi_{ij}^{tot} = \pi_{ij}\chi_{ij}^{obs} + (1 - \pi_{ij})g^{-1}(\lambda_{ij} + g(\chi_{ij}^{obs})). \]

Under the MAR assumption, \( \lambda_{ij} = 0 \), and consequently, \( \chi_{ij}^{obs} = \chi_{ij}^{miss} = \chi_{ij}^{tot} \). This justifies a CC analysis. Values of \( \lambda_{ij} \) away from zero reflect IM; for instance, \( \lambda_{ij} < 0 \) means that participants who leave the study have on average a lower outcome than completers. The choice of the treatment effect measure is separate from the choice of the method used to relate \( \chi_{ij}^{miss} \) to \( \chi_{ij}^{obs} \), that is, different choices may be made for the link functions \( g \) and \( f \). However, ln RoM is likely to be used in conjunction with IMRoM, and MD and SMD with IMDoM, as they are intuitively related.

We need to make assumptions about the parameter \( \lambda_{ij} \), because it is not possible to learn about it from the data. We might choose fixed values, but typically, we are not sure of the value of \( \lambda_{ij} \), so we place a distribution on it with mean \( \mu_{\lambda_{ij}} \) and variance \( \sigma_{\lambda_{ij}}^2 \), written \( \lambda_{ij} \sim N(\mu_{\lambda_{ij}}, \sigma_{\lambda_{ij}}^2) \). In all cases, \( \mu_{\lambda_{ij}} \) describes the most likely MD between the missing and observed outcomes, while \( \sigma_{\lambda_{ij}}^2 \) describes the magnitude...
of uncertainty around this value. Further assumptions can be considered about the means and variances \((\mu_{ij} \text{ and } \sigma^2_{ij})\) and about the similarity of \(\lambda_{ij}\) across treatments, in analogy to the models presented in White et al. [17]. Specifically, we may consider the following:

- Separate \(\lambda_{ij}\) for each arm and study, \(\lambda_{ij} \sim N \left(\mu_{ij}, \sigma^2_{ij}\right)\), where the \(\lambda\)'s are independent both across arms and studies.
- Study-specific \(\lambda: \lambda_{ij} \sim N \left(\mu_{ij}, \sigma^2_{ij}\right)\), where the \(\lambda_i\)'s are independent across studies. The relationship between missing and observed data is different for each study but the same for both arms within a study.
- Correlated across arms. It is assumed that the relation between missing and observed data is different for each study but the same for both arms within a study.
- Correlated across arms. It is assumed that the relation between missing and observed data is different for each study but the same for both arms within a study.

The posterior distribution expresses the probability distribution of \(\lambda_i\) given \(\theta\).

\[
\left(\hat{\lambda}_{IC}, \hat{\lambda}_{IT}\right) \sim N_2 \left(\begin{pmatrix} \mu_{IC} \\ \mu_{IT} \end{pmatrix}, \begin{pmatrix} \sigma^2_{IC} & \rho \sigma_{IC} \sigma_{IT} \\ \rho \sigma_{IC} \sigma_{IT} & \sigma^2_{IT} \end{pmatrix}\right).
\]

2.3. Estimation of the relative treatment effect in a single trial

We are interested in estimating the treatment effect \(\beta\) and its variance in each study, conditional on the prior distribution of \(\lambda\) and the data \(X_{ij, obs}, \chi_{ij, obs}, s_{IT}, s_{IC}, n_{IT}, n_{IC}, m_{IT}\), and \(m_{IC}\). This is therefore a two-stage approach: in the first stage, we adjust effect sizes and their variances in each study for the IM data, while in the second stage, we pool the adjusted effect sizes. We present two methods for estimating \(\beta\) and its uncertainty. We first consider full Bayesian inference and Monte Carlo integration, which we take to be the gold standard. As this approach can be computationally slow, we also present an approximate method based on a Taylor series approximation. Pooling of results across studies is described in Section 3.

2.3.1. Estimation using Monte Carlo integration (parametric bootstrapping). We use the likelihood function for a study \(i, l (\text{data}_{ij})|\theta\), where \(\theta = (\chi_{ij, obs}, \chi_{ij, miss}, \chi_{ij, obs}, \chi_{ij, miss}, \lambda_{IC}, \lambda_{IT})\), and prior information for \(\theta\), which is expressed as a probability density function \(h(\theta)\) before observing any data. The IM parameter \((\lambda_{IT}, \lambda_{IC})\) has the informative prior specified in Section 2.2, whereas \((\chi_{ij, obs, \pi_{IT}}, \chi_{ij, obs, \pi_{IC}})\) have independent noninformative priors. \(\chi_{ij, miss}\) does not appear in \(\theta\) because it is a function of \(\chi_{ij, obs}\) and \(\lambda_{ij}\). The prior beliefs are updated via the likelihood function to yield the posterior distribution for \(\theta\) via Bayes’ theorem

\[
h(\theta|\text{data}_{ij}) = \frac{l(\text{data}_{ij}|\theta)h(\theta)}{\int l(\text{data}_{ij}|\theta)h(\theta)\,d\theta},
\]

where the denominator is a normalizing constant and \(R_{\theta}\) is the set of possible values for \(\theta\). The posterior distribution expresses the probability distribution of \(\theta\), after observing the data. Missingsness parameters \(\lambda\), cannot be informed by the data, and their prior distribution is identical to the posterior. This stresses the importance of the prior distribution to be assigned.

We are interested in estimating \(\beta\), which can be written as \(\beta(\theta)\) using (2) and (4). Specifically, we compute the expected value of \(\beta(\theta)\) over the posterior distribution of \(\theta\), that is, \(E(\beta(\theta)|\text{data}_{ij}) = \int \beta(\theta) h(\theta|\text{data}_{ij})\,d\theta\), and the corresponding variance. The integrals are calculated using Monte Carlo integration using a large number of samples, selected randomly from \(h(\theta|\text{data}_{ij})\). We repeat the following procedure, indexed with \(b\), a large number of times \(B\) for each \(i, j\):

1. Draw \(\chi_{ij, obs}^b\) from \(\chi_{ij, obs}\) \(\sim N\left(\chi_{ij, obs}, \chi_{ij, obs}\right)\).
2. Draw \(\lambda_{ij}^b\) from \(\lambda_{ij}^b \sim N \left(\mu_{ij}, \sigma^2_{ij}\right)\).
3. Set \(\chi_{ij, miss}^b = \chi_{ij, obs}^b + \chi_{ij, obs}^b\) for IMDoM or \(\chi_{ij, miss}^b = \chi_{ij, obs}^b\) for IMRoM.
4. Draw the nonmissing rate \(\pi_{ij}^b\) from \(\pi_{ij}^b \sim N \left(p_{ij}, p_{ij}(1-p_{ij})\right)\).
5. Compute \(\chi_{ij, tot}^b = \chi_{ij, obs}^b + (1 - \pi_{ij}^b) \chi_{ij, miss}^b\).

Steps 1 and 4 use large sample approximations to the t-distribution and the beta distribution, respectively.
Finally, for each $i$, we compute the effect sizes $(\hat{\beta}_i)_b = f \left( (x_{ij}^{\text{total}})_b \right) - f \left( (x_{ij}^{\text{obs}})_b \right)$. The mean and variance of $(\hat{\beta}_i)_b$ across all Monte Carlo samples $b = 1, \ldots, B$ estimate $E(\beta_i|\text{data}_i)$ and $V_i = \text{var}(\beta_i|\text{data}_i)$, respectively.

When $\lambda_{ij}$ and $\lambda_{ic}$ are correlated, we modify step 2 of the aforementioned procedure by sampling $\lambda_i = (\lambda_{ic}^*, \lambda_{ij}^*)'$ from a joint bivariate normal distribution.

2.3.2. Estimation using a Taylor approximation.

2.3.2.1. Estimating the mean effect size. We first estimate $E(x_{ij}^{\text{total}}|\text{data}_i)$. For IMDoM, Equation (4) simplifies to $E(x_{ij}^{\text{total}}|\lambda_{ij}, \text{data}_i) = (1 - p_{ij}) \lambda_{ij} + x_{ij}^{\text{obs}}$, which implies $E(x_{ij}^{\text{total}}|\text{data}_i) = (1 - p_{ij}) \mu_{ij} + x_{ij}^{\text{obs}}$. For IMRoM, Equation (4) simplifies to $E(x_{ij}^{\text{total}}|\lambda_{ij}, \text{data}_i) = (p_{ij} + (1 - p_{ij}) e^{\lambda_{ij}}) x_{ij}^{\text{obs}}$, which implies $E(x_{ij}^{\text{total}}|\text{data}_i) = (p_{ij} + (1 - p_{ij}) E(e^{\lambda_{ij}})) x_{ij}^{\text{obs}}$. Both equations reduce to $E(x_{ij}^{\text{total}}|\text{data}_i) = x_{ij}^{\text{obs}}$ when $\lambda_{ij} = 0$.

We now have the expected value of $\beta_i$, conditional on the data:

$$E(\beta_i|\text{data}_i) = E(f(x_{ij}^{\text{total}}|\text{data}_i)) - E(f(x_{ij}^{\text{total}}|\text{data}_i)) \approx f(E(x_{ij}^{\text{total}}|\text{data}_i)) - f(E(x_{ij}^{\text{total}}|\text{data}_i)).$$

2.3.2.2. Estimating the variance of the mean effect size. The variance of $\beta_i$, conditional on the data can be approximated from the observed data using another Taylor series approximation of Equation 2.

$$V_i = \text{var}(\beta_i|\text{data}_i) \approx \text{var}(x_{ij}^{\text{total}}|\text{data}_i) f'(x_{ij}^{\text{total}})^2 + \text{var}(x_{ij}^{\text{obs}}|\text{data}_i) f'(x_{ij}^{\text{obs}})^2 - 2 \text{cov}(x_{ij}^{\text{total}}|\text{data}_i, x_{ij}^{\text{obs}}|\text{data}_i) f'(x_{ij}^{\text{total}}) f'(x_{ij}^{\text{obs}}).$$

(5)

The covariance term in Equation (5) is only needed if $\lambda_{ic}$ and $\lambda_{ij}$ are correlated.

We show how to evaluate the variances $\text{var}(x_{ij}^{\text{total}}|\text{data}_i)$ and covariance $\text{cov}(x_{ij}^{\text{total}}, x_{ij}^{\text{obs}}|\text{data}_i)$ in the succeeding text. The derivative of the link function $f'(u)$ is $1$ for MD, $1/u$ for ln RoM, and $\frac{1}{\lambda_{ij}}$ for SMD.

To obtain $\text{var}(x_{ij}^{\text{total}}|\text{data}_i)$, we first use a Taylor series approximation of $\text{var}(x_{ij}^{\text{total}}|\lambda_{ij}, \text{data}_i)$ about sample estimators $x_{ij}^{\text{obs}}$ and $p_{ij}$;

$$\text{var}(x_{ij}^{\text{total}}|\lambda_{ij}, \text{data}_i) \approx \text{var}(\pi_{ij}|\lambda_{ij}, \text{data}_i) \left( \frac{d x_{ij}^{\text{total}}}{d \pi_{ij}} \right)^2 + \text{var}(x_{ij}^{\text{obs}}|\lambda_{ij}, \text{data}_i) \left( \frac{d x_{ij}^{\text{total}}}{d x_{ij}^{\text{obs}}} \right)^2$$

In the preceding text, we estimate the variances as

$$\text{var}(\pi_{ij}|\lambda_{ij}, \text{data}_i) = \frac{p_{ij}(1 - p_{ij})}{n_{ij} + m_{ij}}, \quad \text{var}(x_{ij}^{\text{obs}}|\lambda_{ij}, \text{data}_i) = \frac{x_{ij}^{\text{obs}}}{n_{ij}};$$

and the derivatives as

$$\frac{d x_{ij}^{\text{total}}}{d \pi_{ij}} = -\lambda_{ij}, \quad \frac{d x_{ij}^{\text{total}}}{d x_{ij}^{\text{obs}}} = 1$$

when the link function $g$ is the identity and

$$\frac{d x_{ij}^{\text{total}}}{d \pi_{ij}} = (1 - e^{\lambda_{ij}}) x_{ij}^{\text{obs}}; \quad \frac{d x_{ij}^{\text{total}}}{d x_{ij}^{\text{obs}}} = (1 - p_{ij}) e^{\lambda_{ij}} + \pi_{ij};$$

when $g$ is the log function. To allow for the uncertainty in $\lambda_i$, we use $\text{var}(x_{ij}^{\text{total}}|\text{data}_i) = E\left[ \text{var}(x_{ij}^{\text{total}}|\lambda_{ij}, \text{data}_i) \right] + \text{var}\left[ E(x_{ij}^{\text{total}}|\lambda_{ij}, \text{data}_i) \right]$. The first term is approximated by
Expressions for the mean, variance, and covariance are summarized in Table I.

In the Appendix, we provide the variances for all three effect sizes under both IMDoM and IMRoM. In all expressions in the appendix, setting $\mu_{ij} = \sigma_{ij} = 0$ gives $\hat{\beta}_i = x_{iT} - x_{iC}$ and $V_i = \sum_{j=C,T} s_{ij}^2$, which correspond to a CC analysis.

### 3. Pairwise meta-analysis and network meta-analysis models

We showed in Section 2.3 how to estimate $\hat{\beta}_i = E(\beta_i|\text{data}_i)$ and $\hat{V}_i$ for each study accounting for missing data. We now describe how to pool the adjusted estimates $\hat{\beta}_i$. 

---

**Table I.** Estimated means $E(\chi_{ij}^{\text{tot}}|\lambda_{ij}, \text{data}_i)$ and variances $\text{var}(E(\chi_{ij}^{\text{tot}}|\lambda_{ij}, \text{data}_i))$ for arm $j$ in study $i$ under both informative missingness difference of means and informative missingness ratio of means.

| Parameter | Estimated mean in arm $j$ | Variance in arm $j$ | Covariance between arms $j$ and $k$ |
|-----------|--------------------------|---------------------|-----------------------------------|
| $\sigma_{ij}$ | $p_{ij}$ | $p_{ij}(1-p_{ij})/(n_{ij}+m_{ij})$ | 0 |
| $\chi_{ij}^{\text{obs}}$ | $x_{ij}^{\text{obs}}$ | $x_{ij}^{\text{obs}}/n_{ij}$ | 0 |
| $\chi_{ij}^{\text{IMDoM}}$ | $p_{ij}(1-p_{ij})/n_{ij}+m_{ij}$ | $p_{ij}(1-p_{ij})(1-p_{ik})+\sigma_{ij}^2(1-p_{ij})^2+\sigma_{ik}^2(1-p_{ik})^2$ | $\rho_{ij}\sigma_{ij}\sigma_{ik}(1-p_{ij})(1-p_{ik})$ |
| $\chi_{ij}^{\text{IMRoM}}$ | $p_{ij}(1-p_{ij})/n_{ij}+m_{ij}$ | $p_{ij}(1-p_{ij})^2+\sigma_{ij}^2(1-p_{ij})^2+\sigma_{ik}^2(1-p_{ik})^2$ | $\text{cov}(\epsilon_{ij}, \epsilon_{ik})$ |

IMDoM, informative missingness difference of means; IMRoM, informative missingness ratio of means.

var $\left(\chi_{ij}^{\text{tot}}|\lambda_{ij} = \mu_{ij}, \text{data}_i\right)$. The second term is $(1-p_{ij})^2\sigma_{\lambda_{ij}}^2$ for IMDoM and $(1-p_{ij})^2\sigma_{\lambda_{ij}}^2$ var $(\epsilon_{ij})$ for IMRoM.

To obtain cov $\left(\chi_{ij}^{\text{tot}}, \chi_{ij}^{\text{tot}}|\text{data}_i\right)$, the Taylor series approximation gives a term, which is proportional to the correlation $\rho_{\lambda}^2(1-p_{ij})(1-p_{ik})\rho_{ij}\sigma_{ij}\sigma_{ik}$ for IMDoM and $(1-p_{ij})(1-p_{ik})\sigma_{ij}\sigma_{ik}$ cov $(\epsilon_{ij}, \epsilon_{ik})$ for IMRoM. This completes the calculation of Equation 5.

For example, the variance of the MD using IMDoM is

$$V_i \approx \sum_{j=C,T} \left[ p_{ij} \left(1-p_{ij}\right) \left(\mu_{ij}^2 + \sigma_{ij}^2\right) + \frac{s_{ij}^2}{n_{ij}} + \sigma_{ij}^2 \left(1-p_{ij}\right)^2 \right] - 2\rho_{ij}\sigma_{ij}\sigma_{ik}(1-p_{ij})(1-p_{ik})$$

Expressions for the mean, variance, and covariance are summarized in Table I.

In the Appendix, we provide the variances for all three effect sizes under both IMDoM and IMRoM.
3.1. Pairwise meta-analysis

In a random-effects meta-analysis model, for a study $i$, the estimated from Equation 2 effect size $\hat{\beta}_i$ is modelled as

$$\hat{\beta}_i = \mu + \delta_i + \epsilon_i,$$

where $\mu$ is the true effect of active treatment relative to the control treatment, $\delta_i$ is a random effect with $\delta_i \sim N(0, \tau^2)$, where the between-study variance (heterogeneity) $\tau^2$ expresses how effectiveness varies across studies, and $\epsilon_i$ is a sampling error term with $\epsilon_i \sim N(0, V_i)$. In matrix notation, the model can be expressed as

$$\hat{\beta} = X\mu + \delta + \epsilon$$

where $\hat{\beta}$ is a $n$-vector including the adjusted effect sizes for each study, $\epsilon$ is a $n$-vector of normally distributed sampling errors with $\epsilon \sim N(0, \Sigma)$ and known diagonal covariance matrix $\Sigma = \text{diag}(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_n)$. According to Equation 6, each $\hat{\Sigma}_i$ equals the indirect estimate obtained by comparing B and C to a common comparator A ($\mu_{AC} - \mu_{AB}$). In a NMA with $T$ treatments, we select a reference treatment and estimate $T - 1$ basic contrasts that represent the effectiveness of each of the remaining treatments relative to the reference one. The remaining parameters are estimated through the consistency equations. In the aforementioned example, if A is the reference treatment, then $\mu_{AB}$ and $\mu_{AC}$ are the basic contrasts and $\mu_{BC}$ is estimated as $\mu_{BC} = \mu_{AC} - \mu_{AB}$. The NMA model is expressed through Equation 6, but $\mu$ is now a $T - 1$ vector of the basic contrasts, and the design matrix $X$ indicates the treatment comparisons in each study under the consistency assumption [19].

Multi-arm trials complicate this procedure. In a three-arm trial, we estimate two effect sizes, for example, $\beta_i = (\beta_{AB}, \beta_{AC})'$, where $\beta_{ij} = f(\chi_{ij}^{\text{tot}}) - f(\chi_{ij}^{\text{int}})$ with $j = B, C$. We need to estimate the covariance of the two components $\beta_{AB}$ and $\beta_{AC}$, which arises because they include the same control A and because they may have correlated $\lambda_{ij}$. This is handled in the Monte Carlo scheme of Section 2.3.1 by simulating $\chi_{AB}^{\text{tot}}$ and $\chi_{AC}^{\text{tot}}$ from a bivariate normal in step 1. For the Taylor series approximation of Section 2.3.2, the covariance is

$$\text{cov}(\beta_{AB}, \beta_{AC}|\text{data}_i) \approx \text{cov}(\chi_{AB}^{\text{tot}}, \chi_{AC}^{\text{tot}}|\text{data}_i)f'(\chi_{AB}^{\text{tot}})f'(\chi_{AC}^{\text{tot}})\text{cov}(\chi_{AB}^{\text{tot}}, \chi_{AC}^{\text{tot}}|\text{data}_i)f'(\chi_{AB}^{\text{tot}}) + \text{var}(\chi_{AC}^{\text{tot}}|\text{data}_i)f'(\chi_{AC}^{\text{tot}})^2$$

and all these terms have been evaluated in Section 2.3.2.

We have uploaded MATLAB [20] code for the suggested methodology at www.mtm.uoi.gr, and we will soon upload a Stata [21] routine.

© 2014 The Authors. Statistics in Medicine Published by John Wiley & Sons Ltd.

Statist. Med. 2015, 34 721–741
4. Applications to pairwise meta-analysis and network meta-analysis

The suggested methods are illustrated via two examples. The first example synthesizes eight studies comparing the effectiveness of mirtazapine (T) and placebo (C) in patients with major depression [22]. The outcome is the change in depression symptoms measured on a standardized rating scale, the HAMD21 scale. For each study arm, we have the mean change, the standard deviation, the number of patients randomized to this arm, and the number of completers (Table II). We analyze the data using all effect sizes presented in Section 2.1 (MD, SMD, and ln RoM).

We demonstrate how we add uncertainty because of IM data in the fifth study in Table II (MIR 003-021). For the placebo arm, the variance of \( \chi_{5C}^2 \) is \( \frac{(s_{obs}^2)_5C}{n_5C} = 1.34 \) if we assume MAR. Suppose that we use effect measure MD and an IMDoM scenario where \( \lambda \) follows \( N(\mu_{\lambda} = 0, \sigma_{\lambda}^2 = 4) \). Then, the variance of \( \chi_{5C}^2 \) using the Taylor series approximation is

\[
\frac{p_5C(1-p_5C)}{n_5C} + \frac{\lambda_{5C}^2 + \sigma_{\lambda}^2 (1-p_{5C})^2}{n_{5C}^2} = 3.68.
\]

After a similar calculation for the treatment arm, we estimate the treatment effects. If we assume MAR, we obtain \( \hat{\beta}_5 = 3.60 \) with \( \hat{V}_5 = 2.92 \), whereas if we assume \( \lambda_{ij} \sim N(\mu_{\lambda_{ij}} = 0, \sigma_{\lambda_{ij}}^2 = 4) \) independently in the two arms, we obtain \( \hat{\beta}_5 = 3.60 \) with \( \hat{V}_5 = 5.56 \).

The second example is used to illustrate the application of the method to a group of nine antidepressants compared in 12 studies within the context of an NMA [22]. The outcome is change score in depression measured on a standardized rating scale. Data are shown in Table III, and Figure 1 presents the network plot [23]. As studies use two different scales to measure depression, we consider only SMD and ln RoM. These studies were chosen from a large network of antidepressant studies [22], they are not representative of treatment effectiveness and are used for illustrative purposes here. We restricted analysis to this subset.

### Table II. Mirtazapine meta-analysis: mean change in HAMD21 scores, standard deviations, and numbers of completers and noncompleters for the mirtazapine and placebo arms.

| Study          | Placebo | Mirtazapine |
|---------------|---------|-------------|
|               | \( x_{obs} \) | \( s_{obs} \) | \( n \) | \( m \) | \( x_{obs} \) | \( s_{obs} \) | \( n \) | \( m \) |
| Claghorn 1995 | -11.4   | 10.2        | 19   | 26   | -14.5   | 8.8        | 26   | 19   |
| MIR 003-003   | -11.5   | 8.3         | 24   | 21   | -14     | 7.3        | 27   | 18   |
| MIR 003-008   | -11.4   | 8           | 17   | 13   | -13.2   | 8          | 12   | 18   |
| MIR 003-020   | -6.2    | 6.5         | 24   | 19   | -9      | 23         | 21   |
| MIR 003-021   | -17.4   | 5.3         | 21   | 29   | -13.8   | 5.9        | 22   | 28   |
| MIR 003-024   | -11.1   | 9.9         | 27   | 23   | -15.7   | 6.7        | 30   | 20   |
| MIR 84023a    | -11.9   | 8.6         | 33   | 24   | -14.2   | 7.6        | 35   | 25   |
| MIR 84023b    | -11.8   | 8.3         | 48   | 18   | -14.7   | 8.4        | 51   | 13   |

### Table III. Antidepressants network: study name, treatments compared, and observed mean change in depression score, standard deviations, number of completers, and noncompleters for each arm.

| Study             | Comparison                              | Treatment A | Treatment B |
|-------------------|-----------------------------------------|-------------|-------------|
|                   | \( x_{obs} \) | \( s_{obs} \) | \( n \) | \( m \) | \( x_{obs} \) | \( s_{obs} \) | \( n \) | \( m \) |
| Kato 2006         | Fluxovamine versus paroxetine           | 9           | 8.65        | 41   | 8    | 4.6       | 8.65        | 39   | 13   |
| Rossini 2002      | Fluxovamine versus sertaline            | 7.56        | 12.31       | 39   | 1    | 11.27     | 11.33       | 45   | 3    |
| Annseu 1994       | Milnacipran versus fluoxetine           | 17.2        | 11.1        | 74   | 23   | 14.42     | 11.1        | 75   | 18   |
| Yang 2003         | Milnacipran versus sertaline            | 13.5        | 2.1         | 12   | 15   | 13.85     | 1.8         | 15   | 11   |
| Eker 2005         | Reboxetine versus sertaline             | 6.55        | 5.23        | 20   | 5    | 7.76      | 2.89        | 21   | 3    |
| Chen 2001         | Sertaline versus venlafaxine            | 7           | 5           | 45   | 0    | 6         | 5           | 43   | 1    |
| Shelton 2006      | Sertaline versus venlafaxine            | 9.6         | 6.2         | 63   | 19   | 8.4       | 5.4         | 67   | 11   |
| Ekselius 1997     | Citalopram versus sertaline             | 10.4        | 9.12        | 37   | 163  | 11         | 9.12        | 55   | 145  |
| Khanzode 2003     | Citalopram versus fluoxetine            | 15.8        | 3.3         | 30   | 3    | 18.7      | 5.1         | 32   | 2    |
| Diaz/Martinez 1998| Fluoxetine versus venlafaxine           | 8.2         | 7.52        | 55   | 20   | 8.7       | 7.52        | 55   | 15   |
| Amini 2005        | Mirtazapine versus fluoxetine           | 7.4         | 7.52        | 16   | 2    | 10.4      | 7.52        | 15   | 3    |
| Silverston 1999   | Fluoxetine versus venlafaxine           | 12          | 8.65        | 89   | 32   | 11.3      | 8.65        | 91   | 37   |
Figure 1. Network plot for comparison of antidepressants. Nodes are weighted according to the number of studies including the respective interventions. Edges are weighted according to the inverse variance of the direct treatment effect estimates for the respective comparisons.

Figure 2. Comparison of Monte Carlo and Taylor approximation for computing the posterior mean and standard error of the three effect sizes (MR, SMD, and ln RoM) for the MIR 003-21 study assuming a bivariate distribution for IMDoM. Solid line refers to the Taylor series approximation and the dotted line to Monte Carlo sampling.

of the evidence base because observed mean values in each arm before conducting any imputation strategy (i.e., last observation carried forward) are available only for these 12 studies.

Empirical analysis and simulation findings have shown that ln RoM is a reasonable alternative to SMD for pooling continuous outcomes in meta-analysis and it does not involve the estimation of the pooled standard deviation, a quantity not well understood by clinicians [16]. Like SMD, ln RoM is a unit-less measure that can be used irrespective of the units of the trial outcome measures, and it requires the
means in both arms in all trials included in the meta-analysis to have the same sign, which is the case in our examples.

4.1. Comparison of Monte Carlo and Taylor estimation methods in a single study

In Section 2.3, we presented two methods for estimating $E(\beta_i|\text{data}_i)$ and $V_i = \text{var}(\beta_i|\text{data}_i)$, a Monte Carlo method and a Taylor series approximation. Here, we explore how the posterior mean values and standard error for $\beta_i$ vary as uncertainty ($\sigma_{\lambda_i}$) in the IM parameters increases for each study under both methods, considering all three different effect sizes (MD, SMD, and ln RoM). For brevity, we present results only for the fifth study in the mirtazapine example (MIR 003-21 in Table II), which has a high proportion of missing data: 58% in the placebo group and 56% in the mirtazapine group. We assume a bivariate normal distribution for $\lambda_i = (\lambda_{iC}, \lambda_{iT})'$ with mean zero, equal standard deviations in both groups $\sigma_{\lambda_iC} = \sigma_{\lambda_iT} = \sigma_{\lambda_i}$ ranging from 0 to 5 for IMDoM and from 0 to 0.5 for IMRoM, and correlation 0.5. We evaluate the posterior mean and standard error of $\beta_5$ when the effect size is MD, SMD, and ln RoM using (i) Monte Carlo with 10,000 samples and (ii) the Taylor series approximation.

The two methods are compared in Figures 2 and 3 for IMDoM and IMRoM, respectively. We consider Monte Carlo to be the gold standard. In all cases, $V_i$ increases as $\sigma_{\lambda_i}$ increases and represents an increase in uncertainty as we depart from the MAR assumption. The mean and standard error of $\beta_5$ are the same irrespective of the method used (Taylor series approximation or Monte Carlo) for MD and SMD when IMDoM is considered (Figure 2). When ln RoM is considered, using a Taylor series approximation, the expected value is slightly biased upwards and the standard error slightly overestimated for $\sigma_{\lambda_i} > 2$ (Figure 2). When the IMRoM and ln RoM are considered, the standard error is overestimated using the Taylor series approximation for $\sigma_{\lambda_i} > 3$ (Figure 3). This probably occurs because of the small sample properties of ln RoM, which is a nonlinear function of the difference in means between the two arms; ideally, higher-order terms would be included in the Taylor series approximation.

All results in the following applications were estimated using Monte Carlo. The fifth study has the largest missing rate, and a CC analysis gives a 95% confidence interval (CI) for the MD that ranges from

![Figure 3](image-url). Comparison of Monte Carlo and Taylor approximation of computing the posterior mean and standard error of the three effect sizes (MR, SMD, and ln RoM) for the MIR 003-21 study assuming a bivariate distribution for IMRoM. Solid line refers to the Taylor series approximation and the dotted line to Monte Carlo sampling.
0.25 to 6.98 suggesting that placebo is better (Figure 4). If we account for missing outcome data assuming \( \lambda \sim N(0, 1) \), we expect to obtain an estimate with increased uncertainty, and indeed, we obtain (using the corresponding equation from Appendix A) a 95% CI \((-1.05, 8.21)\).

4.2. Pairwise meta-analysis accounting for missingness

We undertake meta-analyses under the MAR assumption and under IM with a common distribution for, independent across arms for MD, SMD and ln RoM. In all cases, we assumed \( \mu_\lambda = 0 \). Results are given in Table IV. In some cases, the summary result of the meta-analysis changes. For instance, in the random-effects model with MD, the summary effect is marginally nonsignificant under MAR, but its CI moves away from zero in the IM analyses as \( \sigma_\lambda \) increases. We see in Figure 4 that the weight of study MIR 003-021 FDA (the fifth study), which favours placebo and has the largest missing rate (56%), is reduced from 17.27% (MAR) to 10.44% when \( \lambda \sim N(0, 4^2) \). Consequently, the summary result moves in favour of mirtazapine as we move away from the MAR assumption.

Figure 5 shows how the three effect sizes (dotted lines/middle line in each plot) change as \( \sigma_\lambda \) increases when IMDoM (panel a) and IMRoM (panel b) are assumed. Solid lines represent the 95% CI as \( \sigma_\lambda \) increases. For MD and SMD, the summary effect decreases, moving away from the line of no effect both for IMDoM and for IMRoM. Precision of the summary effect increases until \( \sigma_\lambda \) is around 0.4 for IMDoM or 0.2 for IMRoM, heterogeneity becomes zero for the three effect sizes and then precision of the summary effect decreases. In general, in a random-effects model, it is expected that increasing \( \sigma_\lambda \) would decrease heterogeneity and increase within-study variance with an unpredictable effect on the overall uncertainty until heterogeneity vanishes; from that point on, it is expected that overall uncertainty will increase.
Table IV. Mirtazapine meta-analysis: pooled intervention effects (95% confidence intervals) and estimated heterogeneity for various model assumptions.

| Assumption about missingness | Fixed effect |  | Random effects |  |
|-----------------------------|-------------|--------------------------------------------------|-------------|
|                             | MD          | SMD | ln RoM  | MD  | SMD | ln RoM |
|                             | –2.05       | –0.30 | 0.12 | –2.34 | –0.29 | 0.19 |
|                             | (–3.51, –0.58) | (–0.49, –0.11) | (0, 0.24) | (–4.67, ≈ 0) | (–0.56, –0.01) | (–0.01, 0.39) |
| MAR/CC analysis             |             |      |       | r² = 6.44 | r² = 0.08 | r² = 0.05 |
| IMDoM                      | –2.41       | –0.33 | 0.15 | –2.42 | –0.33 | 0.18 |
| λy ~ N (0, 2²)              | (–4.12, –0.69) | (–0.55, –0.12) | (0.01, 0.29) | (–4.51, –0.33) | (–0.57, –0.08) | (–0.01, 0.36) |
| IMDoM                      | –2.54       | –0.34 | 0.18 | –2.54 | –0.34 | 0.18 |
| λy ~ N (0, 3²)              | (–4.50, –0.58) | (–0.59, –0.10) | (0, 0.33) | (–4.50, –0.58) | (–0.59, –0.10) | (0, 0.33) |
| IMDoM                      | –2.66       | –0.35 | 0.18 | –2.66 | –0.35 | 0.18 |
| λy ~ N (0, 4²)              | (–4.90, –0.41) | (–0.63, –0.07) | (0, 0.37) | (–4.90, –0.41) | (–0.63, –0.07) | (0, 0.37) |
| IMDoM                      | –2.28       | –0.31 | 0.14 | –2.28 | –0.31 | 0.18 |
| (λi_C / λj_T) ~ N² ((0 0) , (2 4)) | (–3.88, –0.68) | (–0.52, –0.12) | (0.01, 0.27) | (–4.57, –0.18) | (–0.57, –0.05) | (0, 0.37) |
| λy ~ N (0, 0.1²)            | –2.37       | –0.33 | 0.15 | –2.37 | –0.33 | 0.19 |
| IMRoM                      | (–3.97, –0.77) | (–0.53, –0.13) | (0.03, 0.28) | (–4.61, –0.24) | (–0.57, –0.07) | (0, 0.38) |
| λy ~ N (0, 0.2²)            | –2.97       | –0.38 | 0.21 | –2.97 | –0.38 | 0.21 |
| IMRoM                      | (–5.28, –0.66) | (–0.66, –0.09) | (0.03, 0.39) | (–5.28, –0.66) | (–0.66, –0.09) | (0.03, 0.39) |
| λy ~ N (0, 0.3²)            | –2.24       | –0.32 | 0.14 | –2.24 | –0.32 | 0.14 |
| IMRoM                      | (–3.79, –0.69) | (–0.52, –0.12) | (0.02, 0.27) | (–4.63, –0.13) | (–0.57, –0.04) | (0, 0.38) |
| (λi_C / λj_T) ~ N² ((0 0) , (0.1² 0.005 0.005 0.1²)) | (–3.79, –0.69) | (–0.52, –0.12) | (0.02, 0.27) | (–4.63, –0.13) | (–0.57, –0.04) | (0, 0.38) |
| λy ~ N (0, 0.4²)            | 2.38        | –0.31 | 0.19 | 2.38 | –0.31 | 0.19 |

MD, mean difference; SMD, standardized mean difference; ln RoM, logarithm of the ratio of means; MAR, missing at random.
4.3. Accounting for missing data in network meta-analysis

For the antidepressants network, an NMA model is assumed, with a common heterogeneity across treatment comparisons, to estimate relative effectiveness between each pair of antidepressants.

Fluvoxamine is arbitrarily chosen to be the reference treatment. Treatment effects of the eight antidepressants against fluvoxamine, along with their 95% CIs and the estimated overall heterogeneity, are presented in Table V. We analyze SMD under a CC analysis and three assumptions for IMDoM, and ln RoM under a CC analysis and three assumptions for IMRoM.

All studies are small, and relative treatment effects have wide CIs. For SMD, as \( \sigma_\lambda \) increases, CIs for some treatment effects become wider, and there is a shift in the estimated treatment effect for citalopram. For ln RoM, there are no big differences across the different values of \( \sigma_\lambda \).

We assume that indirect estimates are consistent with direct estimates of the same quantity and that we can therefore obtain a reliable ranking of the treatments [19]. Figure 6 shows the cumulative ranking curves when IMDoM is assumed with \( \lambda = 0 \) (solid line) and when \( \lambda \sim N(0, 3^2) \) (dashed line) to produce a visual comparison of the treatments’ effectiveness. The effect size considered is SMD. We also computed the surface under the cumulative ranking curve (SUCRA) [24]. The SUCRA value for each treatment is expressed as a percentage, showing how much effectiveness each treatment achieves compared with an imaginary treatment that is the most effective with absolute certainty. With 12 studies and nine treatments, there is much uncertainty regarding treatment effectiveness. However, the SUCRA values do not seem to change when we move away from the MAR assumption. The only considerable change is for citalopram where the SUCRA value drops from 0.30 when \( \lambda = 0 \) to 0.23 when \( \lambda \sim N(0, 3^2) \). The Ekselius 1997 study comparing citalopram versus sertraline has a high missing rate (more than 70% in both arms). The weight given to the Ekselius study (a study involving citalopram and with a high missing rate) in the computation of the summary estimate reduces substantially as \( \sigma_\lambda \) increases [results not shown].

The direct evidence from the Ekselius study suggests that citalopram and sertraline are equally effective, but this is not supported by the indirect evidence. Sertaline appears to be the second best treatment (SUCRA = 0.67), whereas paroxetine is the second least effective treatment (SUCRA = 0.30).
### Table V. Antidepressants network: summary relative treatment effects for each active treatment versus fluvoxamine (95% confidence intervals) and heterogeneity variances for each effect size under three informative missingness difference of means assumptions for standardized mean difference and three informative missingness ratio of means assumptions for logarithm of the ratio of means using Monte Carlo and assuming a random-effects model.

| Treatment      | CC analysis | \( \lambda \sim N(0, 0.2^2) \) | \( \lambda \sim N(0, 0.3^2) \) | \( \lambda \sim N(0, 4^2) \) | \( \lambda \sim N(0, 0.2^2) \) | \( \lambda \sim N(0, 0.3^2) \) | \( \lambda \sim N(0, 0.1^2) \) | ln RoM |
|----------------|-------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|-------|
| Paroxetine     | -0.51       | (-1.02, 0.00)                   | (-0.97, -0.06)                  | (-1.00, -0.02)                  | (-1.03, 0.02)                  | (-0.98, -0.05)                  | (-1.48, 0.05)                  | (-1.50, 0.06) |
| Sertaline      | 0.32        | (0.18, 0.82)                    | (-0.12, 0.74)                  | (-0.12, 0.76)                  | (-0.13, 0.76)                  | (-0.23, 1.08)                  | (-0.23, 1.08)                  | (-0.23, 1.09) |
| Milnacipran    | 0.46        | (-0.21, 1.14)                   | (-0.12, 1.14)                  | (-0.17, 1.18)                  | (-0.22, 1.22)                  | (-0.24, 1.09)                  | (-0.24, 1.10)                  | (-0.25, 1.14) |
| Fluoxetine     | 0.27        | (-0.33, 0.88)                   | (-0.29, 0.81)                  | (-0.34, 0.83)                  | (-0.38, 0.85)                  | (-0.39, 0.99)                  | (-0.39, 0.99)                  | (-0.39, 0.99) |
| Reboxetine     | 0.03        | (-0.80, 0.86)                   | (-0.77, 0.81)                  | (-0.82, 0.88)                  | (-0.87, 0.92)                  | (-0.77, 0.82)                  | (-0.53, 1.01)                  | (-0.53, 1.01) |
| Venlafaxine    | 0.19        | (-0.39, 0.77)                   | (-0.32, 0.69)                  | (-0.36, 0.70)                  | (-0.37, 0.70)                  | (-0.33, 0.70)                  | (-0.39, 0.96)                  | (-0.39, 0.96) |
| Citalopram     | -0.03       | (-0.67, 0.60)                   | (-0.77, 0.46)                  | (-0.91, 0.45)                  | (-1.00, 0.45)                  | (-0.33, 0.70)                  | (-0.52, 0.85)                  | (-0.53, 0.85) |
| Mitrazapine    | -0.12       | (-1.09, 0.84)                   | (-1.04, 0.76)                  | (-1.09, 0.77)                  | (-1.13, 0.81)                  | (-1.03, 0.77)                  | (-1.00, 0.90)                  | (-1.01, 0.91) |
| \( r^2 \)      | 0.0166      | 0                               | 0                               | 0                               | 0                               | 0                               | 0                               | 0     |

In ln RoM, logarithm of the ratio of means; CC, complete case.
5. Discussion

We propose that a CC analysis be the starting and reference point of any analysis attempting to address missing participant data on continuous outcomes in a meta-analysis of RCTs. However, study results may change if we account for the missing participants. Conducting a sensitivity analysis using various scenarios about the missing participants is recommended to evaluate the robustness of results.

When there is little or no heterogeneity, summary estimates may change considerably after accounting for missing outcome data. A trade-off between increasing sampling variability and decreasing heterogeneity keeps the CI at a comparable width for small values of prior uncertainty in missingness parameters ($\sigma_\lambda$). As $\sigma_\lambda$ increases, CIs for the study estimates increasingly overlap and estimated heterogeneity consequently reduces, leading to increased uncertainty in the summary result. The trade-off between decreasing heterogeneity and increasing sampling variances may be influenced by the method used to estimate heterogeneity or may differ among treatment comparisons. Accounting for missing data may have an impact on the consistency assumption in a NMA. This could be the case if there are no missing data in studies with a specific pair of treatments, but studies including one of these treatments and a common comparator have high missing rates. If this comparator is placebo, it is more likely that high missing rates will be observed in the placebo-controlled trials [2]. An NMA is valid only if the distribution of effect modifiers is the same across all treatment comparisons. Missing outcome data may be seen as an effect modifier, so such a scenario may affect the validity of indirect comparisons.

All the aforementioned methods are based on summary data from individual studies. Several methods including multiple imputation, likelihood techniques, and sensitivity analysis have been suggested to mitigate the impact of missing outcome data in individual RCTs [25]. These methods allow greater flexibility compared with the methods presented in this paper, and further research is needed to adapt our method to handle these more sophisticated analyses.

Throughout the examples, we assigned a common distribution to $\lambda$, for all arms and trials. We also assumed that $\mu_{i\lambda} = 0$. As $\sigma_{i\lambda}$ increases, each $\lambda_{ij}$ assumes a flatter distribution, and the plausible departures from the MAR assumption get larger. Many more scenarios are presented, but we chose to use this
one as we did not have any external information about $\lambda$. In practice, expert opinion should be used to inform the parameters of the joint distribution for $\lambda_{IT}$ and $\lambda_{IC}$. Prior elicitation for the mean values and standard deviation is relatively straightforward. However, eliciting the correlation coefficient in a bivariate distribution from experts is not an easy task. White et al. developed an approach for eliciting the difference between missing and observed outcomes in each trial arm [26, 27]. If individual participant data are available, then the IM parameters might also be informed by analyses using baseline covariates and (for longitudinal RCTs) by analyses using values of the outcome at intermediate times. Such individual participant data also enable improved estimation of the treatment effect: accounting for more variables in the analysis typically makes MAR more plausible [28] so that the values of $\mu_i$ and $\sigma_i$ would move closer to 0.

Our two-stage approach first models the studies separately and then combines their results. A hierarchical model could be constructed that models studies jointly and can be estimated in a single stage in a Bayesian framework, as has been carried out for dichotomous outcomes [9]. A one-stage approach would allow us to assume a common missingness parameter across studies, and unlike our two-stage approach, we might be able to learn about the missingness parameters, although such learning relies on strong assumptions [9].

Appendix A

When the two informative missingness parameters $\lambda_{IT}$ and $\lambda_{IC}$ are not independent, then we need to add an extra term in the variance $V_i$ (Equation 4) to account for their correlation, denoted by $\rho_{\lambda_i}$.

In its general form the variance $V_i = \text{var}(\beta_i | \lambda_i, \text{data})$ will be

$$V_i \approx \text{var}
(\chi_{IT}^{\text{tot}} | \lambda_{IT}, \text{data}) f'(\chi_{IT}^{\text{tot}})^2 + \text{var}(\chi_{IC}^{\text{tot}} | \lambda_{IC}, \text{data}) f'(\chi_{IC}^{\text{tot}})^2 - 2 \rho_{\lambda_i} \sigma_{\lambda_{IT}} \sigma_{\lambda_{IC}} \frac{d\chi_{IT}^{\text{tot}}}{d\lambda_{IT}} \frac{d\chi_{IC}^{\text{tot}}}{d\lambda_{IC}}$$

When the effect size is MD and IMDoM is assumed

$$V_i \approx \sum_{j \in \mathcal{T}} \left[ \frac{p_{ij}(1 - p_{ij})}{n_{ij} + m - ij} \left( \mu_{\chi_{ij}}^2 + \sigma_{\chi_{ij}}^2 \right) + \frac{s_{ij}^2}{n_{ij}} \sigma_{\chi_{ij}}^2 + \frac{s_{ij}^2}{n_{ij}}(1 - p_{ij})^2 \right] - 2 \rho_{\lambda_i} \sigma_{\lambda_{IT}} \sigma_{\lambda_{IC}} (1 - p_{IT})(1 - p_{IC})$$

When the effect size is MD and IMRoM is assumed

$$V_i \approx \sum_{j \in \mathcal{T}} \left[ \frac{p_{ij}(1 - p_{ij})}{n_{ij} + m - ij} \left( 1 - 2e^{\mu_{\chi_{ij}} + \frac{1}{2}\sigma_{\chi_{ij}}^2} + e^{2\mu_{\chi_{ij}} + 2\sigma_{\chi_{ij}}^2} \right) \left( \chi_{ij}^{\text{obs}} \right)^2 + \frac{s_{ij}^2}{n_{ij}} \left( 1 - p_{ij} \right)^2 e^{2\mu_{\chi_{ij}} + 2\sigma_{\chi_{ij}}^2} + 2p_{ij}(1 - p_{ij})e^{\mu_{\chi_{ij}} + \frac{1}{2}\sigma_{\chi_{ij}}^2 + p_{ij}^2 \left( \chi_{ij}^{\text{obs}} \right)^2} + e^{2\mu_{\chi_{ij}} + 2\sigma_{\chi_{ij}}^2} \left( e^{\sigma_{\chi_{ij}}^2} - 1 \right) \left( 1 - p_{ij} \right) \chi_{ij}^{\text{obs}} \right]$$

$$- 2 \rho_{\lambda_i} \sigma_{\lambda_{IT}} \sigma_{\lambda_{IC}} (1 - p_{IT})(1 - p_{IC}) \chi_{IT}^{\text{obs}} \chi_{IC}^{\text{obs}} e^{\mu_{\chi_{IT}} + 0.5\sigma_{\chi_{IT}}^2} e^{\mu_{\chi_{IC}} + 0.5\sigma_{\chi_{IC}}^2}$$

When the effect size is SMD and IMDoM is assumed

$$V_i \approx \frac{1}{S_i^2} \left( \sum_{j \in \mathcal{T}} \left[ \frac{p_{ij}(1 - p_{ij})}{n_{ij} + m - ij} \left( \mu_{\chi_{ij}}^2 + \sigma_{\chi_{ij}}^2 \right) + \frac{s_{ij}^2}{n_{ij}} \sigma_{\chi_{ij}}^2 + \frac{s_{ij}^2}{n_{ij}}(1 - p_{ij})^2 \right] - 2 \rho_{\lambda_i} \sigma_{\lambda_{IT}} \sigma_{\lambda_{IC}} (1 - p_{IT})(1 - p_{IC}) \right)$$

where $S_i = \sqrt{\frac{(n_{ij}-1)s_{ij}^2 + (n_{IC}-1)s_{ij}^2}{n_{ij} + n_{IC} - 2}}$
When the effect size is SMD and IMRoM is assumed

\[ V_i \approx \frac{1}{S_i^2} \left( \sum_{j=C,T} \left[ \frac{p_{ij}(1-p_{ij})}{n_{ij} + m_{ij}} \left( 1 - 2e^{\mu_i + \frac{1}{2}\sigma^2_i} + e^{2\mu_i + 2\sigma^2_i} \right) (\gamma_{ij}^{obs})^2 \right] \right. \\
+ \frac{s_{ij}^2}{n_{ij}} \left( (1-p_{ij})^2 e^{2\mu_i + 2\sigma^2_i} + 2p_{ij}(1-p_{ij})e^{\mu_i + \frac{1}{2}\sigma^2_i} + p_{ij}^2 \right) \\
+ e^{2\mu_i + \sigma^2_i} \left( e^\lambda - 1 \right) \left( (1-p_{ij})\gamma_{ij}^{obs} \right)^2 \left. \right] \\
- 2\rho_i \sigma_{iC} \sigma_{iP} (1-p_{iT})(1-p_{iC}) e^{\mu_i + 0.5\sigma^2_i} e^{\mu_i + 0.5\sigma^2_i} \right) \]

When the effect size is ln RoM and IMDoM is assumed

\[ V_i \approx \left( \sum_{j=C,T} \left[ \gamma_{ij}^{obs} \right] \right)^2 \left( \frac{p_{ij}(1-p_{ij})}{n_{ij} + m_{ij}} \left( \mu_{ij}^2 + \sigma_{ij}^2 \right) + \frac{s_{ij}^2}{n_{ij}} \sigma_{ij}^2 (1-p_{ij})^2 \right) \left( \frac{1}{\chi_{iC}^4} \right) \right. \\
\left. - 2\rho_i \sigma_{iC} \sigma_{iP} (1-p_{iT})(1-p_{iC}) e^{\mu_i + 0.5\sigma^2_i} e^{\mu_i + 0.5\sigma^2_i} \right) \]

When the effect size is ln RoM and IMRoM is assumed

\[ V_i \approx \left( \sum_{j=C,T} \left[ \gamma_{ij}^{obs} \right] \right)^2 \left( \frac{p_{ij}(1-p_{ij})}{n_{ij} + m_{ij}} \left( 1 - 2e^{\mu_i + \frac{1}{2}\sigma^2_i} + e^{2\mu_i + 2\sigma^2_i} \right) (\gamma_{ij}^{obs})^2 \right) \\
+ \frac{s_{ij}^2}{n_{ij}} \left( (1-p_{ij})^2 e^{2\mu_i + 2\sigma^2_i} + 2p_{ij}(1-p_{ij})e^{\mu_i + \frac{1}{2}\sigma^2_i} + p_{ij}^2 \right) \\
+ e^{2\mu_i + \sigma^2_i} \left( e^\lambda - 1 \right) \left( (1-p_{ij})\gamma_{ij}^{obs} \right)^2 \left. \right] \]

Appendix B

Matlab code for addressing missing outcome data in pairwise meta-analysis

```matlab
function [PEtot, stdPEtot, CIPETot, PEtotf, CIPETotf, tauStot, ESTot, wtot, varTot] = metamiss(yt, sdt, nt, mt, mulambdat, sigmalt, yc, sdc, nc, mc, mulambdac, sigmalc, rhol, model, outcome, method)

mnt = mt + (1-mt>0)*0.1;
mnc = mc + (1-mc>0)*0.1;

if method == 'T'
    phatt = nt ./ (nt+mt);
    phact = nc ./ (nc+mc);
end
% Computing the lambda and yobst
% Multiplicative model
ymisst = exp(mulambdat + sigmalt^2/2) .* yt;
ymissc = exp(mulambdac + sigmalc^2/2) .* yc;
else if model == 'A'
end
```

© 2014 The Authors. Statistics in Medicine Published by John Wiley & Sons Ltd.
% Additive model
ymisst=mulambdat+yt;
y missc = mulambdac + yc;
end
end
ytott = phatt.*yt + (1-phatt).*ymisst;
ytotc = phatc.*yc + (1-phatc).*ymissc;

if model == ‘M’
  % Multiplicative models
  f=inline('(p+(1-p)*exp(mulambda))^2*s^2/n+((1-p)*y*exp(mulambda))^2*sigma^2');
  f=inline('p*(1-p)/(n+m)*(-2*exp(mulambda+0.5*sigma^2)+exp(2*mulambda+2*sigma^2))*y^2+s^2/n*((1-p)^2*exp(2*mulambda+2*sigma^2)+2*p*(1-p)*exp(mulambda+0.5*sigma^2)+p^2)+((1-p)*y)^2*exp(2*mulambda+sigma^2)*(exp(sigma^2)-1)');
  covariance = 2*rhol*sigmalt*sigmalc.*(1-nc./(nc+mc)).*(1-nt./(nt+mt)).*exp(mulambdat+0.5*sigmal^2).*exp(mulambdac+0.5*sigmal^2).*yt.*yc;
  for i=1:length(yt)
    stdtotc(i) = sqrt(f(mc(i),mulambdac,nc(i),nc(i)/(nc(i)+mc(i)),sdc(i),sigmalc,yc(i)));
    stdtott(i) = sqrt(f(mt(i),mulambdat,nt(i),nt(i)/(nt(i)+mt(i)),sdt(i),sigmalt,yc(i)));
  end
else if model == ‘A’
  % Additive model
  covariance = 2*rhol*sigmalt*sigmalc.*(1-nc./(nc+mc)).*(1-nt./(nt+mt));
  f=inline('s^2/n+(1-p)^2*sigma^2+(mulambda^2+sigma^2)*p*(1-p)/(n+m)');
  for i=1:length(yt)
    stdtotc(i) = sqrt(f(mc(i),mulambdac,nc(i),nc(i)/(nc(i)+mc(i)),sdc(i),sigmalc));
    stdtott(i) = sqrt(f(mt(i),mulambdat,nt(i),nt(i)/(nt(i)+mt(i)),sdt(i),sigmalt));
  end
end
end
if outcome == ‘M’
  % Mean Difference
  EStot = ytot - ytott;
  var tot = stdtott.^2 + stdtotc.^2 - covariance;
else if outcome == ‘R’
  % Ratio of Means
  EStot = log(abs(ytot))./abs(ytotc);
  var tot = (stdtott./yt).^2 + (stdtotc./yc).^2 - covariance.*1./ytott.*1./ytotc;
else
  % Standardized Mean Difference
  spooled = sqrt(((nt-1).*sdt.^2+(nc-1).*sdc.^2)./(nt+nc-2));
  EStot = (ytott-ytotc)/spooled;
  %var tot = (nt+nc)./(nt.*nc)+EStot.^2./2*(nt+nc-2));
  %var tot = 1./nt+1./nc+EStot.^2./2*(nt+nc-2));
vartot=1./(spooled.^2).*(stdtott.^2+stdtotc.^2-covariance);
end

wtot=1./vartot;
[t Austot]=DLtau(EStot’,wtot’,ones(length(yt),1));
%taustot=0;
tautot=sqrt(taustot);
wtot=1./(vartot+taustot);
PEtot=sum(EStot.*wtot)./sum(wtot);
stdPETot=sqrt(1./sum(wtot));
CIPEtot=[PEtot-norminv(0.975)*stdPETot
PEtot+norminv(0.975)*stdPETot];
PEtotf=sum(EStot.*wtot)./sum(wtot);
stdPETotf=sqrt(1./sum(wtot));
CIPEtotf=[PEtotf-norminv(0.975)*stdPETotf
PEtotf+norminv(0.975)*stdPETotf];
decisiontot=(CIPEtot(2)/CIPEtot(1)>0);
else if method=='B'
B=10000;
lambdac=normrnd(mulambdac,sigmalc,B,length(yc));
lambdat=normrnd(mulambdat+rhol*(lambdacmulambdac),
sigmalt*sqrt(1-rhol^2));
phatt=nt./(nt+mt);
pt=binornd(repmat(nt+mt,B,1),repmat(phatt,B,1))./repmat(nt+mt,B,1);
%pt=repmat(nt./(nt+mt),B,1);
yobst=normrnd(repmat(yt,B,1),repmat(sdt./sqrt(nt),B,1));
phatc=nc./(nc+mc);
pc=binornd(repmat(nc+mc,B,1),repmat(phatc,B,1))./repmat(nc+mc,B,1);
%pc=repmat(nc./(nc+mc),B,1);
yobsc=normrnd(repmat(yc,B,1),repmat(sdc./sqrt(nc),B,1));

if model=='M'
% Multiplicative model
ymisst=exp(lambdat).*yobst;
ymissc=exp(lambdac).*yobsc;
else if model=='A'
% Additive model
ymisst=lambdat+yobst;
ymissc=lambdac+yobsc;
end
end

ytruet=pt.*yobst+(1-pt).*ymisst;
ytruec=pc.*yobsc+(1-pc).*ymissc;

if outcome=='M'
% Mean Difference
ytot=ytruet-ytruec;
else if outcome=='R'
% Ratio of Means
ytot=log(abs(ytrue)/abs(ytruec));
else if outcome=='S'
  % Standardized Mean Difference
  spooled=sqrt(((repmat(nt+mt,B,1).*pt-1).*repmat(sdt,B,1).*2+(repmat(nc+mc,B,1).*pc-1).*repmat(sdc,B,1).*2)./(repmat(nt+mt,B,1).*pt+repmat(nc+mc,B,1).*pc-2));
  ytot=(ytrue-ytruec)./spooled;
end
end

vartot=var(ytot);

wtot=1./vartot;
EStot=mean(ytot);
[taustot]=DLtau(EStot’,wtot’,ones(length(yt),1));

%taustot=0;
tautot=sqrt(taustot);
wttot=1./(vartot+taustot);
%EStot=sum(MDtot.*wttot)./sum(wttot);
%EStot=sum(ROMtot.*wttot)./sum(wttot);
PEtot=sum(EStot.*wttot)./sum(wttot);

stdPEtot=sqrt(1./sum(wttot));
CIPETot=[PEtot-norminv(0.975)*stdPEtot PEtot+norminv(0.975)*stdPEtot];

PEtotf=sum(EStot.*wtot)./sum(wtot);

stdPEtotf=sqrt(1./sum(wtot));
CIPETotf=[PEtotf-norminv(0.975)*stdPEtotf PEtotf+norminv(0.975)*stdPEtotf];

Acknowledgements

Dimitris Mavridis and Georgia Salanti received research funding from the European Research Council (IMMA 260559). Ian White is supported by the Medical Research Council [Unit Programme number U105260558].

References

1. Wahlbeck K, Tuunainen A, Ahokas A, Leucht S. Dropout rates in randomised antipsychotic drug trials. *Psychopharmacology (Berl)* 2001; 155:233–239.
2. Spineli LM, Leucht S, Cipriani A, Higgins JP, Salanti G. The impact of trial characteristics on premature discontinuation of antipsychotics in schizophrenia. *Eur Neuropsychopharmacol* 2013; 23(9):1010–1016.
3. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* 2nd ed. Wiley: New York, 2002.
4. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clinical Trials* 2012; 9(4):396–407.
5. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338.
6. Mavridis D, Chaimani A, Efthimiou O, Leucht S, Salanti G. Addressing missing outcome data in meta-analysis. Statistics in practice. *Evidence Based Mental Health* 2014; 17:85–89.
7. Gamble C, Hollis S. Uncertainty method improved on best-worst case analysis in a binary meta-analysis. *Journal of Clinical Epidemiology* 2005; 58:579–588.
8. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; 88:125–134.
9. White IR, Welton NJ, Wood AM, Ades AE, Higgins JPT. Allowing for uncertainty due to missing data in meta-analysis—Part 2 : hierarchical models. *Statistics in Medicine* 2008; 27:728–745.

10. Higgins JPT, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials* 2008; 5:225–239.

11. Magder LS. Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios and risk differences. *Controlled Clinical Trials* 2003; 24:411–421.

12. Spineli LM, Higgins JP, Cipriani A, Leucht S, Salanti G. Evaluating the impact of imputations for missing participant outcome data in a network meta-analysis. *Clinical Trials* 2013; 10(3):378–388.

13. Schafer JL, Graham JW. Missing data : our view of the state of the art. *Psychological Methods* 2002; 7:147–177.

14. Ebrahim S, Akl EA, Mustafa RA, Sun X, Walter SD, Heels-Ansdell D, Alonso-Coello P, Johnston BC, Guyatt GH. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. *Journal of Clinical Epidemiology* 2013; 66:1014–1021.

15. Higgins JPT, Deeks JJ, Altman DG. Chapter 16 : special topics in statistics. In *Cochrane Handbook for Systematic Reviews of Interventions*, Higgins JPT, Green S (eds). John Wiley & Sons, 2008.

16. Friedrich JO, Adhikari NKJ, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis : a simulation study. *BMCMedical Research Methodology* 2008; 8.

17. White IR, Higgins JPT, Wood AM. Allowing for uncertainty due to missing data in meta-analysis-Part 1 : two-stage methods. *Statistics in Medicine* 2008; 27:711–727.

18. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis : many names, many benefits, many concerns for the next generation evidence synthesis school. *Research Synthesis Method* 2012; 3:80–97.

19. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008; 17:279–301.

20. MATLAB The MathWorks. Natick, Massachusetts, United States.

21. Corp Stata. *Stata Statistical Software: Release 12, version College Station*. StataCorp LP: TX, 2011.

22. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins J, Churchill R et al. Comparative efficacy and acceptability of 12 new-generation antidepressants : a multiple-treatments meta-analysis. *Lancet* 2009; 373:746–758.

23. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One* 2013; 8:e76654.

24. Salanti G, Ades A, Ioannidis J. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* 2011; 64:163–171.

25. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials* 2004; 1:368–376.

26. Jackson D, White IR, Leese M. How much can we learn about missing data? An exploration of a clinical trial in psychiatry. *Journal of the Royal Statistical Society, series A* 2010; 173:593–612.

27. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. *Clinical Trials* 2007; 4:125–139.

28. Groenwold RH, Donders AR, Roes KC, Harrell FE Jr, Moons KG. Dealing with missing outcome data in randomized trials and observational studies. *American Journal of Epidemiology* 2012; 175(3):210–217.