Addressing missing outcome data in meta-analysis

Dimitris Mavridis,^{1,2} Anna Chaimani,¹ Orestis Efthimiou,¹ Stefan Leucht,^{3,4} Georgia Salanti¹

¹Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece

²Department of Primary Education, University of Ioannina, Ioannina, Greece

³Department of Psychiatry and Psychotherapy, Technische Universität München, München, Germany

⁴Department of Psychiatry, University of Oxford, Oxford, UK

Correspondence to Dimitris Mavridis, dimi.mavridis@googlemail.com

ABSTRACT

Objective Missing outcome data are a common problem in clinical trials and systematic reviews, as it compromises inferences by reducing precision and potentially biasing the results. Systematic reviewers often assume that the missing outcome problem has been resolved at the trial level. However, in many clinical trials a complete case analysis or suboptimal imputation techniques are employed and the problem is accumulated in a quantitative synthesis of trials via meta-analysis. The risk of bias due to missing data depends on the missingness mechanism. Most statistical analyses assume missing data to be missing at random, which is an unverifiable assumption. The aim of this paper is to present methods used to account for missing outcome data in a systematic review and meta-analysis.

Methods The following methods to handle missing outcome data are presented: (1) complete cases analysis, (2) imputation methods from observed data, (3) best/worst case scenarios, (4) uncertainty interval for the summary estimate and (5) a statistical model that makes assumption about how treatment effects in missing data are connected to those in observed data. Examples are used to illustrate all the methods presented.

Results Different methods yield different results. A complete case analysis leads to imprecise and potentially biased results. The best-case/worst-case scenarios give unrealistic estimates, while the uncertainty interval produces very conservative results. Imputation methods that replace missing data with values from the observed data do not properly account for the uncertainty introduced by the unobserved data and tend to underestimate SEs. Employing a statistical model that links treatment effects in missing and observed data, unlike the other methods, reduces the weight assigned to studies with large missing rates.

Conclusions Unlike clinical trials, in systematic reviews and meta-analyses we cannot adapt pre-emptive methods to account for missing outcome data. There are statistical techniques implemented in commercial software (eg, STATA) that quantify the departure from the missing at random assumption and adjust results appropriately. A sensitivity analysis with increasingly stringent assumptions on how parameters in the unobserved and observed data are related is a sensible way to evaluate robustness of results.

INTRODUCTION

The term attrition is widely used in the clinical trials literature to refer to situations where outcome data are not available for some participants. Missing data may invalidate results from clinical trials by reducing precision and, under certain circumstances, yielding biased results. Missing outcome data are an important and common problem in mental health trials as dropout rate may exceed 50% for certain conditions.¹ The Cochrane Collaboration regards incomplete outcome data as a major factor affecting the credibility of a study and requires systematic reviewers to assess the level of bias in all included trials via the Cochrane Risk of Bias tool.^{2,3} The ideal solution is to avoid missing data altogether and the National Research Council suggested ideas for limiting the possibility of missing data in the design of clinical trials.⁴ Systematic reviews and meta-analysis are retrospective by their own nature and preventive measures for the avoidance of missing outcome data cannot be used.

The intention-to-treat (ITT) principle is widely accepted as the most appropriate way to analyse data in randomised controlled trials (RCT).^{5,6} The ITT principle requires analysing all participants in the group they were originally randomised irrespective of the treatment they actually received. The Cochrane Handbook suggests employing an ITT as the least biased way to estimate intervention effects from randomised trials.⁷ However, in order to include in the analysis participants whose outcomes are unknown, one needs to employ an imputation technique and make assumptions about missing data, which may affect the reliability and robustness of study findings.

Missing data mechanisms

There are several reasons why data may be missing and not all of them introduce bias. The risk of bias due to missing data depends on the missing data mechanism which describes how propensity for missing data depends on the participant's characteristics and outcomes. Missing data mechanisms can be classified as follows:

I. Missing completely at random (MCAR)

The probability of a missing outcome is the same for all participants and does not depend on any participant characteristic (eg, if a participant misses some appointments due to scheduling difficulties). The MCAR assumption means that the group of participants who provided data is a random sample of the total population of participants, but this is often unrealistic in practice.

II. Missing at random (MAR)

The propensity for missingness is related to participants' characteristics, but the probability of a missing outcome is not related to the outcome itself. For instance, suppose that primary school children are randomised to different interventions aiming to reduce school-related anxiety measured on a symptom severity scale. Younger children are less likely to provide data because they have a harder time understanding the items of the symptom severity scale. In the study, the proportion of young children and missing rates are expected to be comparable across interventions. The MAR assumption implies that outcomes for the younger children who dropped out are expected to be similar to outcomes for the younger children who completed the study.⁴ Under the MAR assumption, missingness does not depend on the actual outcome, although it is associated with some participant's or setting's characteristics. This term MAR is often confusing and sometimes misunderstood as MCAR. Statistical analyses usually start with MAR assumption, and, if it is true, analysis of completers only can provide an unbiased estimate of the relative treatment effect.⁸ However, the MAR assumption is formally untestable in meta-analyses as we usually have aggregated data and not enough information about those who dropped out; hence, we should consider a different approach for dealing with missing data. By contrast, it is possible to explore the MAR assumption using auxiliary data within an individual trial, for example, if baseline disease severity predicts missingness it is sensible to assume that final disease severity would predict missingness as well and the data may not be MAR.⁹

III. Missing not at random (MNAR) or informatively missing (IM)
Even accounting for all the available observed information, the probability that an observation is missing still depends on the unseen observations themselves. Participants may dropout for reasons that are associated with the actual effect of the intervention. In schizophrenia trials, for example, placebo arms show larger dropout rate than patients treated with antipsychotics because of placebo's lack of efficacy. Analysis of the participants who completed the study under MNAR would provide a biased estimate of the relative treatment effect. When missing data are MCAR or MAR they are termed ignorable. A MNAR mechanism is termed non-ignorable.

We use a hypothetical example to illustrate differences between the three categories. Consider an RCT with 200 participants randomised equally (1:1) to the experimental or control group (table 1). We assume that the true response rate is 33.3% in the control group and 50% in the experimental group so that the estimated OR should be 2. In the MCAR scenario, 10% of the participants dropped out because they missed the appointment. In the MAR scenario, young people dropped out because summer started and they left for vacation. In both these scenarios the missing rate across treatment groups is the same because groups are expected to have participants with similar baseline characteristics (eg, age). Therefore, the probability of dropping out is the same in both the groups. In the MNAR scenario, 40% of the participants who did not see any improvement dropped the study. The number of those not improved is larger in the control group; hence, missing rate is larger in the control group. Reasons for dropping out are related to the intervention received and more specifically to the actual outcome of the study. Missingness shed bias in the results that favours the control group because successes in completers remained unchanged in both groups but the number of completers is now smaller in the control group. We get unbiased results both in the MCAR and MAR scenarios, but to ignore missing data in the MNAR scenario may give biased results. Bias increases with sample size, hence large studies with data that are MNAR may give seriously biased results.

METHODS

Methods to address missing data may power results and increase precision. Systematic reviewers are encouraged to collect information about the amount of missing outcome data and the techniques used to estimate treatment effects in each trial. There is a plethora of both ad hoc and more sophisticated approaches for handling missing outcome data at the meta-analysis level.¹⁰

Imputations of missing outcome data in individual studies and their meta-analysis

Trials often employ methods to impute values for missing outcomes. Two very common approaches include (1) replacing missing values with the mean value of the participants who provided data (simple imputation) and (2) replacing missing values with the last observed value. The latter method is called last observation carried forward (LOCF) and is routinely used in mental health trials. The LOCF approach has been criticised for producing biased results, as conditions in mental health field are rarely stable and usually involve progressive conditions.^{11 12} Hence, it is not sensible to assume that participants who dropped out at intermediate steps would have remained stable until the end of the study (especially when data are MNAR the LOCF cannot be defended because participants could have left the study for reasons associated with their unobserved outcome data eg, they got worse). Simple imputation of missing values using the two aforementioned methods usually underestimates the SE for the outcome because it fails to account for the fact that missing values are imputed rather than observed. Multiple imputations replace missing values by a number of different plausible imputed values and subsequently combine them to get parameter estimates.¹³ Plausible imputed values are sampled from their predictive distributions given the observed data (eg, using regression models). The SE of the treatment effect is estimated by the *within* variance of each imputed dataset as well as from the variance *between* the imputed datasets. A naïve or unprincipled imputation method may create more problems than it solves by introducing bias both in parameter estimates and their SE.⁸ When data are MAR, multiple imputation may correct for bias but it will still yield biased results when MNAR holds. To give unbiased results, the variable(s) that are predictive of missing data should be included in the imputation model.

Published study reports typically present results together for fully observed and imputed outcomes. Consequently, a meta-analysis is not given much choice but to synthesise study outcomes as reported in trials, even when the imputation technique is inappropriate. In some cases, trial reports present the outcomes for completers only as well as the results from the merged sample of observed and imputed outcomes.

Synthesis of studies with missing outcome data

It is often the case that studies report results from the participants who provided the outcome of interest and they only report the number of participants for which the outcome is unknown. From a meta-analysis perspective several synthesis options exist and are outlined below.

Table 1 Hypothetical example of missing data mechanisms applied to a randomised trial with 100 participants in each arm. We assume that the true response rate is 33.3% in the control group and 50% in the experimental group so that the estimated OR should be 2

	Missing data mechanism					
	MCAR		MAR		MNAR	
	control group	experimental group	control group	experimental group	control group	experimental group
Completers	90	90	60	60	73	80
Non-completers	10	10	40	40	27	20
Successes in completers	30	45	20	30	30	45
Reason for dropping out	10% of participants missed their appointment		Young people (40% of the total sample) dropped out because summer started and wanted to leave for vacation		40% of participants who did not see any improvement dropped out (roughly 67% in the control group and 50% in the experimental group did not see any improvement, hence on average the 40% who dropped out refers to 27 and 20 participants, respectively)	
Estimated OR (using completers)	2.00		2.00		1.84	
	95% CI (1.10 to 3.65)		95% CI (0.96 to 4.18)		95% CI (0.97 to 3.50)	

Complete cases meta-analysis

This is usually the reference approach in many meta-analyses. From each study, only individuals whose outcome is known are included. If the MCAR assumption holds then a complete cases meta-analysis will give unbiased results and the only consequence of missing data is the loss in power. The larger the missing rate, the less reliable the results of this analysis when the data are MNAR.¹⁰

Best-case and worst-case scenarios

A typical simple imputation technique for dichotomous outcomes includes the best-case and worst-case scenarios.¹⁴ The best-case scenario assumes that all missing participants have a favourable outcome in the experimental group and poor outcome in the control group; the converse is assumed for the worst-case scenario. These two extremes are typically used as a sensitivity analysis and may produce unrealistic results in practice, especially if missing rates are high.

Uncertainty intervals

Gamble and Hollis¹⁴ suggested that studies for which there is a big discrepancy between best-case and worst-case scenarios should be down-weighted. Best-worst case scenarios give an interval, called *uncertainty interval*, for the treatment effect that includes all uncertainty due to missing data.¹⁴ A pseudo SE is estimated from this for each study and is subsequently used to downweigh studies. A summary estimate is computed using the revised weights from the uncertainty intervals. This method has low power when there is a large amount of missing data.¹⁴

Using a statistical model that relates missing data to observed data

For the case of missing binary outcomes, White and colleagues presented a meta-analysis model where the degree of departure from the MAR assumption is quantified by the informative missingness OR (IMOR).¹⁵ IMOR describes the relationship between the unknown odds of the outcome among missing participants and the known odds among observed participants.^{15 16} By relating missing treatment effects

to the observed ones it is possible to adjust treatment effects within each study. The adjusted treatment effects are then summarised via meta-analysis. Higgins *et al*¹⁶ suggested a sensitivity analysis for the IMOR approach with risks imputed for the missing data over a plausible range of values. Expert opinion can be used to elicit information on how the risk in the missing participants is related to that of the observed participants and use that information to adjust treatment effects. Experts may be asked how larger/smaller is the risk in the missing participants compared with that of the observed participants for each treatment group and study. Giving a range of plausible value for the ratio or difference of risks may help quantify uncertainty around IMOR. The IMOR approach is implemented in STATA¹⁷ in the *metamiss* command¹⁸ (<http://www.mrc-bsu.cam.ac.uk/software/stata-software/>). The IMOR approach can incorporate the best-case/worst-case scenarios as special cases and has several advantages. It does not aim to estimate the missing outcomes but aims at making valid inference on the summary treatment effects.¹⁹ It also accounts for the uncertainty induced by missing outcome data, unlike the naïve approaches that consider the imputed values as if they were fully observed. An obvious downside of this approach is its complexity compared with the naïve approaches as it requires involvement of a knowledgeable statistician and expert opinion.

RESULTS

The approaches described above are illustrated via an example of a systematic review of all registration studies which compared the effectiveness of an atypical antipsychotic (amisulpride) with a typical one (haloperidol or flupenthixol) in schizophrenia.^{20–24} The condition of the participants was measured with the Brief Psychiatric Rating Scale (BPRS). The primary outcome is a 50% reduction in BPRS between baseline and week 6. We carried out a meta-analysis and results are reported here.

In figure 1 we present the ORs and the weights used in a random-effects meta-analysis reporting corresponding estimates for each one of the five studies, using complete cases meta-analysis (black

Figure 1 OR for short-term improvement (response) in participants with schizophrenia allocated to amisulpride or typical antipsychotics. Complete cases meta-analysis (†), best case scenario (‡), worst case scenario (§), synthesis of studies where missing outcomes have been imputed with last observation carried forward (¶) and informative missingness OR (#) models (see text for details).

lines), best case scenario (blue), worst case scenario (red), synthesis of studies where missing outcomes have been imputed with LOCF (green) and IMOR model (brown). For the IMOR approach we assumed that the odds in the missing participants is on average the same to that of the observed participants (IMOR=1) but with some uncertainty expressed by the fact that there is 70% chance that the IMOR varies from 0.5 to 2. The differences between the results depend heavily on the rate of missing participants. Studies by Colonna and Carriere, for example, have the lowest dropout rates (about 10%) and hence the ORs slightly change across the different methods. By contrast, studies with much higher dropout rates (about 30%) present results that are very conflicting when different assumptions are made for the missing outcome data. Best-case/worst-case scenario analysis gives extreme results, whereas the LOCF approach reduces erroneously uncertainty, because it considers imputed data as if they were truly observed. The IMOR approach, as opposed to the other ones, drastically reduces the weights assigned to studies having large missing rates, while the relative weights increase for studies with lower missingness.

In figure 2 we present the pooled meta-analytical results for each approach. The LOCF method provides the most precise estimate, while the Gamble and Hollis uncertainty interval is very wide and probably there is not enough power to detect a non-zero effect. The differences in significance between the complete cases analysis and the IMOR model is attributable to the fact that IMOR reduces the weight of Möller 1997 study²³ which has a large missing rate and an OR that favours (though not significantly) the typical antipsychotics. Although the IMOR approach increases within-study variability, it may also result in a decrease in between-study variability and uncertainty around the summary estimate does not necessarily increase when a random-effects model is assumed.

DISCUSSION

Meta-analysts typically do not have access to the reasons for missing outcome data and consequently the MAR assumption cannot be tested empirically. A sensitivity analysis is the only viable way to evaluate the effect of different scenarios for the missing data mechanism.¹³ Missing data should not be ignored and systematic reviewers should not take for granted that the problem has been appropriately handled in the trial level. Naïve imputation techniques such as imputing the mean value, LOCF and the best-case and worst-case scenarios, though widely used, may produce biased results and underestimated SEs. The uncertainty interval has low power.

Figure 2 Meta-analytical results for the ORs of short-term improvement in schizophrenia (response) for amisulpride versus other typical antipsychotics: various alternative approaches for handling missing outcome data.

Assuming a statistical model that relates treatment effect in missing data to those of the observed data is helpful to consider how robust results are to departures from the MAR assumption. Clinical expertise may inform such a model or a sensitivity analysis can be employed assuming increasingly more stringent scenarios.

The Cochrane Handbook⁷ urges systematic reviewers to

- contact the original investigators to request missing data;
- state explicitly the assumptions of any methods used to cope with missing values;
- perform sensitivity analysis to assess how robust results are;
- address the potential impact of missing data on the findings of the review.

In most trials, most of all in mental health trials, missing data are likely not to be missing at random. Researchers should document the reasons why data are missing and collect data on auxiliary variables that may be predictive of both the outcome and the probability of dropping out.

Competing interests DM and GS received research funding from the European Research Council (IMMA 260559), AC and OE received funding from Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF)—Research Funding Program: ARISTEIA. Investing in knowledge society through the European Social Fund.

doi:10.1136/eb-2014-101900

REFERENCES

- Wahlbeck K, Tuunainen A, Ahokas A, *et al*. Dropout rates in randomised antipsychotic drug trials. *Psychopharmacology (Berl)* 2001;**155**:230–3.
- Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Fellow JPHSSV, Director SGF, eds. *Cochrane handbook for systematic reviews of interventions* [Internet]. John Wiley & Sons, Ltd, 2008:187–241 [cited 2014 May 6]. <http://onlinelibrary.wiley.com/doi/10.1002/9780470712184.ch8/summary>
- Higgins JPT, Altman DG, Gotzsche PC, *et al*. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928.
- Little RJ, D'Agostino R, Cohen ML, *et al*. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012;**367**:1355–60.
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;**319**:670–4.
- Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol* 1992;**21**:837–41.
- Higgins JP, Deeks JJ, Altman DG. Special topics in statistics. In: Fellow JPHSSV, Director SGF, eds. *Cochrane handbook for systematic reviews of interventions* [Internet]. John Wiley & Sons, Ltd, 2008:481–529 [cited 2014 May 5]. <http://onlinelibrary.wiley.com/doi/10.1002/9780470712184.ch16/summary>
- Little RJA, Rubin DB. *Statistical analysis with missing data*. New York, NY, USA: John Wiley & Sons, Inc, 1986.
- White IR, Welton NJ, Wood AM, *et al*. Allowing for uncertainty due to missing data in meta-analysis—part 2: hierarchical models. *Stat Med* 2008;**27**:728–45.
- Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials Lond Engl* 2004;**1**:368–76.
- Leucht S, Engel RR, Bäuml J, *et al*. Is the superior efficacy of new generation antipsychotics an artifact of LOCF? *Schizophr Bull* 2007;**33**:183–91.
- Mallinckrodt CH, Watkin JG, Molenberghs G, *et al*. Choice of the primary analysis in longitudinal clinical trials. *Pharm Stat* 2004;**3**:161–9.
- Sterne JAC, White IR, Carlin JB, *et al*. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;**338**:b2393.
- Gamble C, Hollis S. Uncertainty method improved on best-worst case analysis in a binary meta-analysis. *J Clin Epidemiol* 2005;**58**:579–88.
- White IR, Higgins JPT, Wood AM. Allowing for uncertainty due to missing data in meta-analysis—part 1: two-stage methods. *Stat Med* 2008;**27**:711–27.
- Higgins JP, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clin Trials Lond Engl* 2008;**5**:225–39.
- Stata Statistical Software: Release 13. College Station, TX: StataCorp LP, 2013.
- White IR, Higgins JPT. Meta-analysis with missing data. *Stata J* 2009;**9**:57–69.
- Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 2002;**7**:147–77.
- Puech A, Fleuret O, Rein W. Amisulpride, and atypical antipsychotic, in the treatment of acute episodes of schizophrenia: a dose-ranging study vs. haloperidol. The Amisulpride Study Group. *Acta Psychiatr Scand* 1998;**98**:65–72.

21. **Colonna L**, Saleem P, Dondey-Nouvel L, *et al*. Long-term safety and efficacy of amisulpride in subchronic or chronic schizophrenia. Amisulpride Study Group. *Int Clin Psychopharmacol* 2000;**15**:13–22.
22. **Carrière P**, Bonhomme D, Lempérière T. Amisulpride has a superior benefit/risk profile to haloperidol in schizophrenia: results of a multicentre, double-blind study (the Amisulpride Study Group). *Eur Psychiatry J Assoc Eur Psychiatr* 2000;**15**: 321–9.
23. **Möller HJ**, Boyer P, Fleurot O, *et al*. Improvement of acute exacerbations of schizophrenia with amisulpride: a comparison with haloperidol. PROD-ASLP Study Group. *Psychopharmacology (Berl)* 1997;**132**:396–401.
24. **Wetzel H**, Gründer G, Hillert A, *et al*. Amisulpride versus flupentixol in schizophrenia with predominantly positive symptomatology—a double-blind controlled study comparing a selective D2-like antagonist to a mixed D1-/D2-like antagonist. The Amisulpride Study Group. *Psychopharmacology (Berl)* 1998;**137**:223–32.