



OPEN ACCESS



Potential impact of missing outcome data on treatment effects in systematic reviews: imputation study

Lara A Kahale,^{1,2} Assem M Khamis,³ Batoul Diab,¹ Yaping Chang,⁴ Luciane Cruz Lopes,⁵ Arnav Agarwal,^{4,6} Ling Li,⁷ Reem A Mustafa,^{4,8} Serge Koujanian,⁹ Reem Waziry,¹⁰ Jason W Busse,^{4,11,12,13} Abeer Dakik,¹ Holger J Schünemann,⁴ Lotty Hooft,² Rob JPM Scholten,² Gordon H Guyatt,^{4,14} Elie A Akl¹⁴

For numbered affiliations see end of the article.

Correspondence to: E A Akl ea32@aub.edu.lb (or @elie__akl on Twitter: ORCID 0000-0002-3444-8618)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;370:m2898 <http://dx.doi.org/10.1136/bmj.m2898>

Accepted: 30 June 2020

ABSTRACT

OBJECTIVE

To assess the risk of bias associated with missing outcome data in systematic reviews.

DESIGN

Imputation study.

SETTING

Systematic reviews.

POPULATION

100 systematic reviews that included a group level meta-analysis with a statistically significant effect on a patient important dichotomous efficacy outcome.

MAIN OUTCOME MEASURES

Median percentage change in the relative effect estimate when applying each of the following assumption (four commonly discussed but implausible assumptions (best case scenario, none had the event, all had the event, and worst case scenario) and four plausible assumptions for missing data based on the informative missingness odds ratio (IMOR) approach (IMOR 1.5 (least stringent), IMOR 2, IMOR 3, IMOR 5 (most stringent)); percentage of meta-analyses that crossed the threshold of the null effect for each method; and percentage of meta-analyses that qualitatively changed direction of effect for each method. Sensitivity analyses based on the eight different methods of handling missing data were conducted.

RESULTS

100 systematic reviews with 653 randomised controlled trials were included. When applying the implausible but commonly discussed assumptions,

the median change in the relative effect estimate varied from 0% to 30.4%. The percentage of meta-analyses crossing the threshold of the null effect varied from 1% (best case scenario) to 60% (worst case scenario), and 26% changed direction with the worst case scenario. When applying the plausible assumptions, the median percentage change in relative effect estimate varied from 1.4% to 7.0%. The percentage of meta-analyses crossing the threshold of the null effect varied from 6% (IMOR 1.5) to 22% (IMOR 5) of meta-analyses, and 2% changed direction with the most stringent (IMOR 5).

CONCLUSION

Even when applying plausible assumptions to the outcomes of participants with definite missing data, the average change in pooled relative effect estimate is substantive, and almost a quarter (22%) of meta-analyses crossed the threshold of the null effect. Systematic review authors should present the potential impact of missing outcome data on their effect estimates and use this to inform their overall GRADE (grading of recommendations assessment, development, and evaluation) ratings of risk of bias and their interpretation of the results.

Introduction

Despite efforts to reduce the occurrence of missing outcome data in clinical trials,¹ this is still common. Across six methodological surveys the percentage of randomised controlled trials with missing outcome data ranged from 63% to 100%,²⁻⁷ and the average proportion of participants with missing outcome data among trials reporting missing data ranged from 6% to 24%.²⁻⁸ Among 235 randomised controlled trials with statistically significant results published in leading medical journals, one in three lost statistical significance when making plausible assumptions about the outcomes of participants with missing data.² Another study comparing different approaches to modelling binary outcomes with missing data in an alcohol clinical trial yielded different results with various amount of bias depending on the approach and missing data scenario.⁹

The extent of missing outcome data in randomised controlled trials contributes to the risk of bias of meta-analyses involving those trials. Additional factors that might bias the results of meta-analysis include the methods used by contributing randomised controlled trials to handle missing data and the transparency of reporting missing data (for each arm and follow-up time point). To explore the impact on

WHAT IS ALREADY KNOWN ON THIS TOPIC

Missing data on the outcomes of participants in randomised controlled trials might introduce bias in systematic reviews

To assess that risk of bias, the GRADE (grading of recommendations assessment, development, and evaluation) working group recommends challenging the robustness of the meta-analysis effect estimate by conducting sensitivity analyses with different methods of handling missing data

WHAT THIS STUDY ADDS

Even when applying plausible assumptions to the outcomes of participants with definite missing data, the average change in pooled relative effect estimate is substantive, whereas almost a quarter (22%) of meta-analyses crossed the threshold of the null effect

Systematic review authors should present the potential impact of missing outcome data on their effect estimates and use this to inform their overall GRADE ratings of risk of bias and their interpretation of the results

risk of bias, the GRADE (grading of recommendations assessment, development, and evaluation) working group recommends conducting sensitivity analyses using assumptions regarding the outcomes of patients with missing outcome data.¹⁰ A recent empirical comparison of bayesian modelling strategies for missing binary outcome data in network meta-analysis found that using implausible assumptions could negatively affect network meta-analysis estimates.¹¹ No methodological study has yet assessed the impact of different assumptions about missing data on the robustness of the pooled relative effect in a large representative sample of published systematic reviews of pairwise comparisons.

One challenge when handling missing data is the lack of clarity in trial reports on whether participants have missing outcome data.¹² We recently published guidance for authors of systematic reviews on how to identify and classify participants with missing outcome data in the trial reports.¹³ The Cochrane Handbook acknowledges that attempts to deal with missing data in systematic reviews are often hampered by incomplete reporting of missing outcome data by trialists.¹⁴ A recent methodological survey among 638 randomised controlled trials reported that the median percentage of participants with unclear follow-up status was 9.7% (interquartile range 4.1-19.9%),⁸ and that when authors explicitly reported not following-up participants, almost half did not specify how they handled missing data in their analysis.

We assessed risk of bias associated with missing outcome data in systematic reviews with two interventions by quantifying the change in effect estimate when applying different methods of handling missing outcome data; examining how these methods alter crossing the threshold of the null effect of pooled effect estimates; qualitatively changing the direction of effect; and exploring the potential effect on heterogeneity of each of these approaches.⁴

Methods

Design

This study is part of a larger project examining methodological problems related to missing data in systematic reviews and randomised controlled trials.¹⁵ Our published protocol includes detailed information on the definitions, eligibility criteria, search strategy, selection process, data abstraction, and data analysis.¹⁵ In the appendix (supplementary table 1), we present deviations from the protocol and the rationale for these deviations.

We defined missing data as outcome data for trial participants that are not available to authors of systematic reviews (from the published randomised controlled trial reports or personal contact with the trial authors).

In the current study, we collected a random sample of 50 Cochrane and 50 non-Cochrane systematic reviews published in 2012 that reported a group level meta-analysis of a patient important dichotomous efficacy outcome, with a statistically significant effect

estimate (the meta-analysis of interest).¹⁶ We used the term original pooled relative effect to refer to the result of the meta-analysis as reported by the systematic review authors. For the individual trials included in the meta-analyses of interest,⁸ we abstracted detailed information relevant to the statistical analysis and missing data and conducted sensitivity meta-analyses based on nine different methods of handling missing data. Our outcomes were the median change of effect estimates across meta-analyses; the percentage of meta-analyses that crossed the threshold of the null effect with each of these methods; the percentage of meta-analyses that changed direction of effect; and the change in heterogeneity associated with each of the methods.

Identifying participants with missing data

Since publication of our protocol, we published a guidance for authors of systematic reviews on how to identify and classify participants with missing outcome data in the trial reports depending on how trial authors report on those categories and handle them in their analyses (table 1).¹³ The guidance includes a taxonomy of categories of trial participants who might have missing data, along with a description of those categories (supplementary table 2). The categorisation reflects the wording used in trial reports—that is, the presentation faced by systematic review authors. We used this guidance to judge the outcome data missingness of categories of participants who might have missing data (ie, whether they have definite, potential, or no missing data).

Data abstraction

Eleven reviewers trained in health research methodology abstracted the data independently and in duplicate. The reviewers met regularly with a core team (EAA, LAK, BD, and AD) to discuss progress and challenges and develop solutions. We used a pilot tested standardised data abstraction form hosted in an electronic data capture tool, REDCap.¹⁷ All reviewers underwent calibration exercises before data abstraction to improve reliability, and a senior investigator served as a third independent reviewer for resolving disagreements.

From each eligible meta-analysis we abstracted the original (published) pooled relative effect—that is, pooled relative effect measure (relative risk or odds ratio) and the associated 95% confidence interval, the analysis model used (random effects or fixed effect), and the statistical method used for pooling data (eg, Mantel-Haenszel, Peto).

For each study arm in the randomised controlled trials, we abstracted the numbers of participants randomised, events, participants with definite missing data (according to the suggested guidance on identifying participants with missing data¹³), and participants with potentially missing data.

Data analysis

SPSS statistical software, version 21.0 was used to conduct a descriptive analysis of study characteristics

Table 1 | Judging missingness of outcome data on the basis of reporting and handling of categories of participants who might have missing data

Categories of participants	Judgment of outcome data missingness
Participants explicitly reported as followed-up, participants who died during the trial, participants belonging to centres that were excluded	Definitely not missing data
Participants explicitly reported as not followed-up, and participants with unclear follow-up status and who were excluded from the denominator of the analysis (ie, complete case analysis) or were included in the denominator of the analysis and their outcomes were explicitly stated to be imputed	Definite missing data
Participants with unclear follow-up status (eg, included in the denominator of the analysis and their outcomes were not explicitly stated to be imputed)	Potential missing data
Participants with definite or potential missing data	Total possible missing data

of eligible systematic reviews and the associated randomised controlled trials.¹⁸ For categorical variables, we report frequencies and percentages, and for continuous variables that were not normally distributed, we use medians and interquartile ranges.

To explore the robustness of pooled effect estimates reported by systematic reviews, we conducted several sensitivity analyses based on nine different methods of handling missing data, using Stata software release 12¹⁹ (see table 2):

- Complete case analysis^{10 20}
- Four implausible but commonly discussed assumptions: best case scenario, none of the participants with missing data had the outcome, all participants with missing data had the outcome, and worst case scenario.
- Four plausible assumptions using increasingly stringent values for informative missing odds ratio (IMOR) in the intervention arm.^{21 22} IMOR describes the ratio of odds of the outcome among participants with missing data to the odds of the outcome among observed participants. In other words, to obtain the odds among participants with missing data, the odds is multiplied among the observed participants with a stringent value (ie, 1.5, 2, 3, and 5). We chose an upper limit of 5 as it represents the highest ratio reported in the literature. One study used a community tracker to evaluate the incidence of death among participants in scale-up programmes of antiretroviral treatment in Africa who were lost to follow-up.^{23 24} The study found the mortality rate to be five times higher in patients lost to follow-up compared with patients who were followed up. We did not use an IMOR of 1 because it provides the same effect estimate as the complete case analysis.

Naïve single imputation methods falsely increase precision by treating imputed outcome data as if they were observed, whereas IMORs (or pattern mixture models in general), by imputing risks, increase uncertainty within a trial to deal with the fact that data have been imputed. Imputing events consists of including participants with missing data in the denominator and making assumptions about their outcomes in the numerator. This approach might lead to imputing several events as if they were fully observed, leading to possibly problematic narrower confidence

intervals than would be the case. To correct for this, methodologists have developed methods that account for uncertainty associated with imputing missing observations using statistical approaches.^{21 22 27} The command “metamiss” in Stata integrates uncertainty within its calculation^{25 26} (see statistical notes in appendix for further details).

We did not consider the nature of outcomes (positive versus negative) during the conduct of the sensitivity analyses best case and worst case scenarios. Instead, we focused in the sensitivity analyses on challenging the effect estimates against the null value when the best case scenario shifts the effect estimate away from the null value of 1 and the worst case scenario shifts the effect estimate towards the null value of 1 (irrespective of whether the outcome is positive or negative).

Our analytical approach was executed in one command (metamiss^{25 26}) of Stata release 12¹⁹ for each sensitivity analysis. Firstly, we recalculated each meta-analysis of interest, for all 100 systematic reviews, using each method to deal with missing outcome data to generate different sensitivity analysis pooled effects along with the corresponding 95% confidence intervals. We used the same relative effect measure (relative risk or odds ratio), the same analysis model (random effects or fixed effect), and the same statistical method (eg, Mantel-Haenszel, Peto) as the original meta-analysis of interest. Secondly, across all included meta-analyses and for each method, we explored the impact of the revised meta-analysis for several outcomes:

Change in relative effect estimate

To quantify the percentage change in relative effect estimate between the sensitivity analysis pooled effect estimate (assumption) and the sensitivity analysis pooled effect estimate (complete case analysis), we applied the formula (fig 1; see statistical notes in appendix for further details).

We calculated specifically the percentage of meta-analyses with change of relative effect estimate (by direction) between the sensitivity analysis pooled effect estimate (assumption) and the sensitivity analysis pooled effect estimate (complete case analysis) and the median and interquartile range for the change in relative effect estimate (stratified by direction of change).

This relative change could be an increase or a reduction in effect. For example, a relative increase

Table 2 | List and description of different methods of handling missing outcome data

Method of handling missing data	Handling participants with missing data in the numerator and denominator			
	Intervention arm		Control arm	
Complete case analysis	Numerator excluded	Denominator excluded	Numerator excluded	Denominator excluded
Implausible but commonly discussed assumptions:				
Best case scenario*	Assumed that all had a favourable outcome	Denominator included	Assumed that all had an unfavourable outcome	Denominator included
None of the participants with missing data had the outcome	Assumed that none had the outcome	Denominator included	Assumed that none had the outcome	Denominator included
All participants with missing data had the outcome	Assumed that all had the outcome	Denominator included	Assumed that all had the outcome	Denominator included
Worst case scenario†	Assumed that all had an unfavourable outcome	Denominator included	Assumed that all had a favourable outcome	Denominator included
Plausible assumptions‡:				
IMOR 1.5	IMOR 1.5§	Denominator included	IMOR 1	Denominator included
IMOR 2	IMOR 2§	Denominator included	IMOR 1	Denominator included
IMOR 3	IMOR 3§	Denominator included	IMOR 1	Denominator included
IMOR 5	IMOR 5§	Denominator included	IMOR 1	Denominator included

IMOR=informative missing odds ratio
 *When applying best case scenario, it was ensured it challenges the relative effect by shifting it away from the null value of no effect (see statistical notes in appendix section 3).
 †When applying worst case scenario, it was ensured it challenges the relative effect by shifting it closer to the null value of no effect (see statistical notes in appendix section 3).
 ‡The "metamiss" command^{25 26} was used to implement the IMOR assumptions in Stata.
 §These calculations are applied when the relative effect is less than 1. When a relative effect is greater than 1, the values for the IMOR are flipped between the intervention and control arm whereby it is 1 for the intervention arm. For example, when an original relative effect is greater than 1, the IMOR value for the intervention arm would be 1 and that of the control arm would be 5.

in relative risk of 25% for the sensitivity analysis pooled effect estimate (worst case scenario) over the sensitivity analysis pooled effect estimate (complete case analysis) implies that for a relative risk of 0.8 with complete case analysis, the relative risk for the worst case scenario would be 1, and for a relative risk of 1.6 with complete case analysis, the relative risk for the worst case scenario would be 1.2.

Crossing the threshold of the null effect

For the analysis of the percentage of meta-analyses for which the sensitivity analysis pooled relative effect (assumption) crossed the threshold of the null effect compared with the sensitivity analysis pooled relative effect (complete case analysis), we restricted the sample to the meta-analyses that did not cross the threshold of the null effect under the complete case analysis method.

Changing direction of pooled relative effect

In the analysis of the percentage of meta-analyses for which the sensitivity analysis pooled relative effect (assumption) changed direction compared with the sensitivity analysis pooled relative effect (complete case analysis), the direction could change from favouring the intervention to favouring the control or vice versa. For this analysis, we restricted the sample to the meta-analyses that crossed the threshold of the null effect under the complete case analysis method.

We reproduced the analyses for outcomes when crossing the threshold of the null effect and when changing direction, comparing the sensitivity analysis pooled relative effect to the original pooled relative effect. In addition, we conducted all the analyses twice: first considering participants with definite missing outcome data, then considering participants with total possible missing outcome data (see table 1). In addition, we explored how heterogeneity varies across the different methods.

Patient and public involvement

As this project concerns methodology, no patients or public were involved.

Results

Study characteristics of included meta-analyses

We previously reported on the details of the 100 eligible systematic reviews¹⁶ and the 653 randomised controlled trials they considered.⁸ Table 3 summarises the characteristics of the 100 included systematic reviews and the corresponding meta-analyses. Most reported on a drug related outcome (61%) and a non-active control (55%), assessed a morbidity outcome (56%), reported an unfavourable outcome (73%), used the risk ratio (61%), and applied a fixed effect analysis model (57%) and Mantel-Haenszel statistical methods (77%). The median number of randomised controlled trials per meta-analysis was 6 (interquartile range 3-8).

$$\frac{\text{Sensitivity analysis pooled relative effect (assumption)} - \text{sensitivity analysis pooled relative effect (CCA)}}{\text{Sensitivity analysis pooled relative effect (CCA)}} \times 100$$

Fig 1 | Formula to quantify percentage change in relative effect. CCA=complete case analysis

Eight meta-analyses included randomised controlled trials that reported no missing data.

Missing data in randomised controlled trials

The results of 400 of the 653 randomised controlled trials (63%) mentioned at least one of the predefined categories of participants who might have missing outcome data. Among those 400 trials, the total median percentage of participants with definite missing outcome data was 5.8% (interquartile range 2.2-14.8%); 3.8% (0-12%) in the intervention arm and 3.4% (0-12%) in the control arm. The median percentage of participants with potential missing outcome data was 9.7% (4.1-19.9%) and with total possible missing data was 11.7% (5.6-23.7%). Only three trials provided a reason for missingness (eg, missing at random).

Change in relative effect estimate

Figure 2 shows the change in the relative effect estimate between the sensitivity analysis pooled effect estimate (assumption) and the sensitivity analysis pooled effect estimate (complete case analysis), when considering participants with definite missing data.

For the four implausible but commonly discussed assumptions, the percentage of meta-analyses with an increased relative effect estimate (shifted away from the null value of 1) was 91% for the best case scenario assumption, 25% for the assumption that none of the participants with missing data had the event, and 17% for the assumption that all participants with missing data had the event. The median increase in the relative effect estimate ranged from 0% for the worst case scenario assumption to 18.9% (interquartile range 6.8-38.9%) for the best case scenario assumption. The percentage of meta-analyses with reduced relative effect estimates (shifted closer towards the null value of 1) was 90% for the worst case scenario assumption, 38% for the assumption that none of the participants with missing data had the event, and 75% for the assumption that all participants with missing data had the event. The median reduction in the relative effect estimate ranged from 0% for the best case scenario assumption to 30.4% (interquartile range 10.5-77.5%) for the worst case scenario assumption.

For the plausible assumptions based on the IMOR, the percentage of meta-analyses with an increased relative effect estimate was 85% for the least stringent assumption (IMOR 1.5) and 88% for the most stringent assumption (IMOR 5). The median reduction in relative effect estimate ranged from 1.4% (0.6-3.9%) for the IMOR 1.5 assumption to 7.0% (2.7-18.2%) for the IMOR 5 assumption. The appendix presents the details of the percentage change in the relative effect estimate when considering participants with total possible missing data (supplementary results C), and stratified by whether the estimate is less than or greater than 1 under the complete case analysis, using either definite missing data or total possible missing data (supplementary results D table 3).

Crossing the threshold of the null effect and change of direction

Of the 100 meta-analyses, 87 did not cross the threshold of the null effect under the complete case analysis method. Figure 3 shows the number of meta-analyses that crossed the threshold of the null effect when comparing the sensitivity analysis pooled relative effect (assumption) with the sensitivity analysis pooled relative effect (complete case analysis) for each assumption and considering participants with definite missing data. For the four implausible but commonly discussed assumptions, the percentage of meta-analyses that crossed the threshold of the null effect ranged from 1% (best case scenario and none of the participants with missing data had the event) to 18% (all participants with missing data had the event) to 60% (worst case scenario). For the plausible assumptions based on IMOR, the percentage of meta-analyses that crossed the threshold of the null effect ranged from 6% (least stringent assumption IMOR 1.5) to 22% (most stringent assumption IMOR 5).

The percentage of meta-analyses that changed direction with the two extreme assumptions was

Table 3 | General characteristics of included systematic reviews and the meta-analyses (n=100). Values are numbers (percentages) unless stated otherwise

Characteristics	Estimate
Median (interquartile range) No of randomised controlled trials in each meta-analysis	6 (3-8)
Type of intervention:	
Drug	61 (61.0)
Surgery or invasive procedure	24 (24.0)
Other	15 (15.0)
Type of control:	
Active: drug	21 (21.0)
Active: surgery or invasive procedure	18 (18.0)
Non-active: no intervention, standard of care, placebo, or sham	55 (55.0)
Other	6 (6.0)
Outcome category:	
Mortality	21 (21.0)
Morbidity	56 (56.0)
Patient reported outcomes	23 (23.0)
Favourability of outcome*:	
Favourable	27 (27.0)
Unfavourable	73 (73.0)
Mean (SD) duration of outcome follow-up (months)	12.5 (23.1)
Effect measures reported:	
Risk ratio	61 (61.0)
Odds ratio	39 (39.0)
Analysis model:	
Random effects model	43 (43.0)
Fixed effect model	57 (57.0)
Statistical methods:	
Mantel-Haenszel	77 (77.0)
Inverse variance	4 (4.0)
Peto	7 (7.0)
Other	7 (7.0)
Not reported	5 (5.0)
Reported handling method:	
Complete case analysis	2 (2.0)
Assuming no participants with missing data had the event	3 (3.0)
Assuming all participants with missing data had the event	2 (2.0)
Not reported	93 (93.0)

*Whether outcome was negative (eg, mortality) or positive (eg, survival).

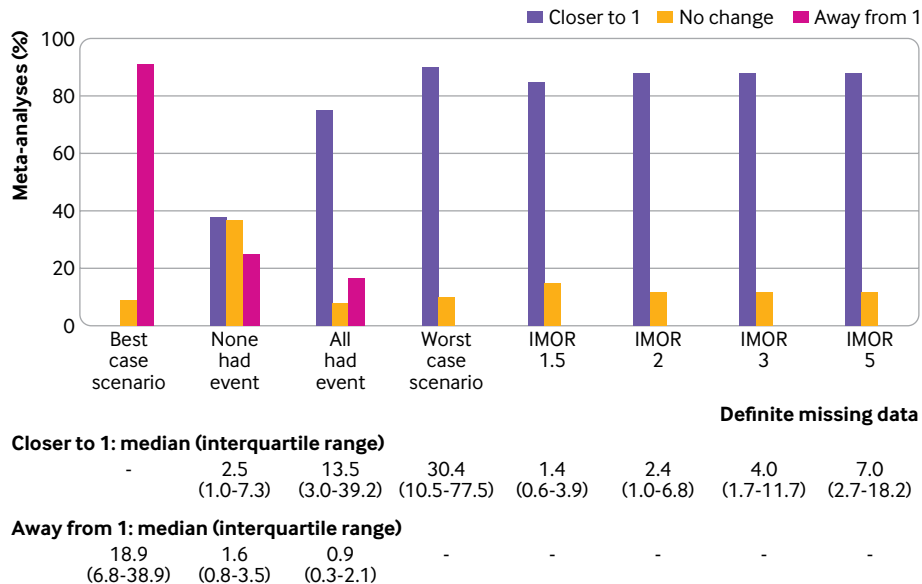


Fig 2 | Change in relative effect estimate (by direction) between the sensitivity analysis pooled effect estimate (assumption) and the sensitivity analysis pooled percentage of meta-analyses effect estimate (complete case analysis) when considering participants with definite missing data. Coloured bars represent the percentage of meta-analyses with change in relative effect estimate (by direction). Numerical values represent the median (interquartile range) for increase and decrease in relative effect estimate (n=100, respectively. IMOR=informative missing odds ratio

26% for the worst case scenario and 2% for the most stringent assumption IMOR 5.

The appendix presents the results of meta-analyses that crossed the threshold of the null effect¹ and changed direction when participants with total possible missing data were considered² (supplementary results A), and when the sensitivity analysis pooled relative effect (assumption) was compared with the original pooled relative effect (supplementary results B).

Change in heterogeneity

Figure 4 shows the change in heterogeneity across the different methods of handling missing data when considering participants with definite missing data. The median I^2 for the original pooled relative effect was 0% (interquartile range 0-42%) and for the complete case analysis was 3.6% (0-47%). For the four implausible but commonly discussed assumptions, the median I^2 ranged from 0.4% (none of the participants with missing data had the event) to 20.3% (all participants with missing data had the event) to 29% (best case scenario) to 58.6% (worst case scenario). For the plausible assumptions based on IMOR, the percentage of meta-analyses that crossed the threshold of the null effect ranged from 2.4% (least stringent assumption IMOR 1.5) to 4.4% (most stringent assumption IMOR 5).

Discussion

In the current study, we quantified the change in the effect estimate when applying different methods of handling missing data to the outcomes of participants with definite missing data. When applying plausible assumptions to the outcomes of participants with definite missing data, the median change in relative

effect estimate was as high as 7.0% (interquartile range 2.7-18.2%). When applying implausible but commonly discussed assumptions, the median change in the relative effect estimate was as large as 30.4% (10.5-77.5%).

We also examined how different methods of handling missing data alter crossing the threshold of the null effect of pooled effect estimates of dichotomous outcomes. Even when applying plausible assumptions to the outcomes of participants with definite missing data, almost a quarter (22%) of meta-analyses crossed the threshold of the null effect. When applying implausible but commonly discussed assumptions, the percentage of systematic reviews that crossed the threshold of the null effect was as high as 60% with the worst case scenario.

Strengths and limitations of this study

In this study we assessed the effect of using different assumptions (both commonly discussed and more plausible) on a large number of published meta-analyses of pairwise comparisons targeting a wide range of clinical areas. Strengths of our study include a detailed approach to assessing participants with missing data, and accounting for participants with potential missing data in our analyses. We used two statistical approaches to assess the risk of bias associated with missing outcomes: change in effect estimate and crossing the threshold of the null effect. Although the approach of crossing the threshold of the null effect has been criticised as the basis for decision making,²⁸ we used it to assess the robustness of the meta-analysis effect estimate (ie, when conducting sensitivity meta-analyses using different methods of handling missing data). We are confident that if results

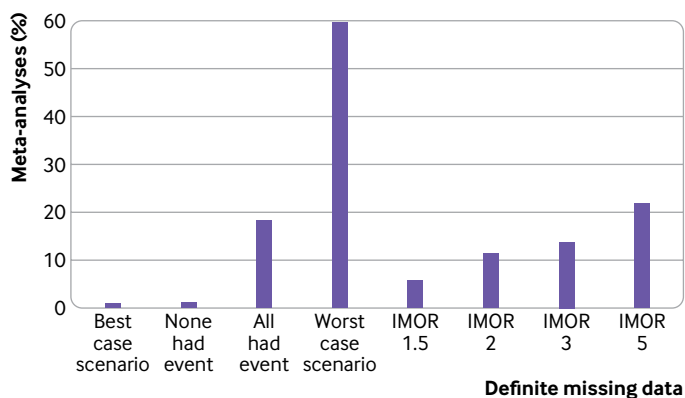


Fig 3 | Results of meta-analyses that crossed the threshold of null effect when considering participants with definite missing data and comparing the sensitivity analysis pooled relative effect (assumption) with the sensitivity analysis pooled relative effect (complete case analysis) (n=87 systematic reviews that did not cross the threshold of null effect under the complete case analysis method). IMOR=informative missing odds ratio

cross the threshold of the null effect, the certainty in evidence should be rated down owing to risk of bias. The change in effect estimate approach has its own limitation in interpretation (cut-off for topic specific minimally important difference that would vary across a wide range of topics and outcomes).

A limitation of our study is that we considered only dichotomous outcome data; methods for handling missing continuous data are different and our findings might not be generalisable to systematic reviews of continuous outcomes.^{29 30} Our sample consisted of systematic reviews that were published in 2012 and these might not reflect more current reviews; however, recent surveys have found that the reporting, handling, and assessment of risk of bias in relation to missing data has not improved over the date of our search.^{7 31-33}

For further confirmation that current practice is unlikely to have changed, we assessed the reporting and handling of missing data in a sample of recently published systematic reviews. To make our selection of a sample of systematic reviews reproducible, we ordered systematic reviews published from January

2020 in the chronological order of their publication and selected the first 15 that met our eligibility criteria. Then we applied the same methodology stated in the methods section to assess the reporting and handling of missing data in the 15 eligible meta-analyses. As with the 2012 sample of systematic reviews reported on here, most of the systematic reviews published in 2020 did not explicitly plan (in the Methods section) to consider any category of missing data (60%), did not report (in the Results section) data for any category of missing data (67%), did not report the proportion of missing data for each trial and for each arm (67%), did not explicitly state a specific analytical method for handling missing data (80%), and did not provide a justification for the analytical method used to handle missing data (100%).

Another limitation was our focus on systematic reviews with statistically significant results. Although this prevented us from assessing the change in effect estimate for meta-analyses with non-statistically significant results, it allowed us to focus on reviews that are more likely to influence clinical practice.

We acknowledge the possible superiority of using the random IMOR over fixed IMOR. However, the variance or uncertainty increases as a function of the proportion of missingness and the variance of the IMOR parameter.^{1 2 5} Hence, uncertainty increases even with fixed IMORs although to a smaller extent. Fully introducing uncertainty to IMORs by using random IMOR (although it was not feasible for this study) would exaggerate our results. Specifically, we would expect a further increase in the proportion of meta-analyses in which the confidence interval would cross the line of no effect and thus lose statistical significance, larger within study variance, and further downgrading of the certainty of the evidence. Hence, the switch to random IMOR would only reinforce the inferences from our work that we already make.

Interpretation of findings

Experience and acceptance for using plausible assumptions is growing as a result of face validity.^{11 34} The advantage of the IMOR approach is that it

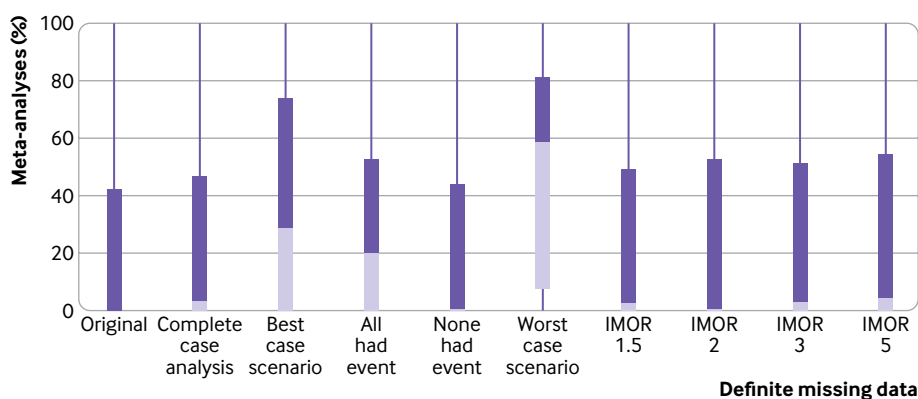


Fig 4 | Change in heterogeneity (I^2) across different methods of handling missing data. IMOR=informative missing odds ratio

provides a tool for review authors to challenge the robustness of effect estimates by applying increasingly stringent assumptions.^{2–5} This approach allows the review authors to choose the IMOR value or range of values based on their clinical judgment or expert opinion. Only if the effect estimate is robust to the most conservative and plausible scenario can it be concluded that the evidence is at low risk of bias from missing data without additional sensitivity analyses. The selection of the assumptions about missing data should be determined a priori with researchers blinded to the extent of missing data in the included trials to avoid data dredging. Thus we should rate down for missing data when it might change inferences. Inferences could be that treatment is of benefit (versus not of benefit) or that treatment achieves an important benefit. That would require a threshold of minimal clinically important difference and a check to see if the threshold was crossed initially and then after accounting for minimal clinically important difference. We could not do that unless we knew the minimal clinically important difference.

Almost a quarter (22%) of meta-analyses crossed the threshold of the null effect when we used a conservative approach to test robustness (ie, applying plausible assumptions to the outcomes of participants with definite missing data). When using the same conservative approach, up to a quarter of meta-analyses had a change of at least 18% in their relative effect estimates (based on the 75th centile for IMOR 5, see median and interquartile range of IMOR 5 in fig 2). These findings mean that a substantive percentage of meta-analyses is at serious risk of bias associated with missing outcome data. Findings such as these should lead systematic review authors to rate down the certainty of evidence for risk of bias. Our results highlight the importance of minimising missing data for clinical trials^{35–36} by better reporting and handling of missing data.^{8–16–37}

With the two assumptions that all participants with missing data had the event and none of the participants with missing data had the event, the size and direction of the confidence intervals are unpredictable; sometimes the effect estimate is shifted closer to 1 and sometimes away from 1. Thus, these assumptions are not helpful for challenging the robustness of the effect estimate. By design, the effect estimate using the best case scenario assumption is shifted in the opposite direction of challenging the robustness and should not be used for that purpose. The worst case scenario consistently challenges the robustness of the effect estimate as it is shifted towards the null effect, but the implausibility of its underlying assumptions makes it a poor choice for sensitivity analyses. If the effect estimate is robust to the worst case scenario it can be concluded that the evidence is at low risk of bias from missing data, without proceeding with further sensitivity analyses.

As for the change in heterogeneity, since with the implausible but common assumptions, the size and direction of the confidence intervals are unpredictable,

consequently the change in I^2 was observed to be unpredictable. As for the plausible assumptions, since uncertainty is taken into account, confidence intervals tend to be wider and more overlapping, leading to low I^2 values. Future studies might need to study whether a certain change in I^2 in a sensitivity analysis is worth rating down the certainty of evidence as a result of inconsistency.

Conclusion and implications

The findings of this study show the potential impact of missing data on the results of systematic reviews. This has implications when both the risk of bias associated with missing outcome data is assessed and the extent of missing outcome data in clinical trials needs to be reduced.

Systematic review authors should present the potential impact of missing outcome data on their effect estimates and, when these suggest lack of robustness of the results, rate down the certainty of evidence for risk of bias. For practical purposes, authors of systematic reviews might wish to use statistical software that allows running assumptions about missing data (eg, Stata). As for users of the medical literature, a rule of thumb on how to judge risk of bias associated with missing outcome data at the trial level is needed. This would account for factors such as percentage of missing data in each study arm, the ratio of missing data to event rate for each study arm (ie, the higher the ratio, the larger the change), fragility of statistical significance (ie, borderline significance), the magnitude of the effect estimate (ie, the larger the effect estimate, the smaller the change), and the duration of follow-up (ie, the longer the duration of follow-up, the higher the percentage of missing data).

We acknowledge that assessing the impact of missing data with crossing the threshold of the null effect might be insensitive. Thus, when using this approach and the threshold of null effect is crossed, then rate down the certainty for risk of bias associated with missing data. If threshold of null effect is not crossed, it might be valuable to then evaluate the change in effect estimate to assess whether the relative effect goes from an important to an unimportant effect. If the latter happens, then rate down the certainty for risk of bias associated with missing data. However, the judgment of whether the change in effect estimate is clinically significant requires using minimal clinically important difference, which varies by clinical question. Thus, it would be ideal to reproduce this study in specific subjects of medical science with clearly defined minimal clinically important differences and using random IMOR instead of fixed IMOR.

Future research could also validate some of the findings of this study. For example, this study could be reproduced using individual participant data meta-analyses and findings compared with the current study. In addition, individual participant data meta-analyses would allow testing other imputation methods, such as multiple imputations.

AUTHOR AFFILIATIONS

¹Department of Internal Medicine, American University of Beirut Medical Center, PO Box 11-0236, Riad-El-Solh Beirut 1107 2020, Beirut, Lebanon

²Cochrane Netherlands and Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

³Wolfson Palliative Care Research Centre, Hull York Medical School, University of Hull, Hull, UK

⁴Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

⁵Pharmaceutical Sciences Post Graduate Course, University of Sorocaba, UNISO, Sorocaba, Sao Paulo, Brazil

⁶Department of Medicine, University of Toronto, Toronto, ON, Canada

⁷Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu, China

⁸Departments of Medicine and Biomedical and Health Informatics, University of Missouri-Kansas City, Kansas City, MO, USA

⁹Department of Evaluative Clinical Sciences, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

¹⁰Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, USA

¹¹Department of Anaesthesia, McMaster University, Hamilton, ON, Canada

¹²The Michael G DeGroot Institute for Pain Research and Care, McMaster University, Hamilton, ON, Canada

¹³Canadian Veterans Chronic Pain Centre of Excellence, Hamilton, ON, Canada

¹⁴Department of Medicine, McMaster University, Hamilton, ON, Canada

Contributors: LAK, AMK, GG, HJS, and EAA conceived and designed the paper. LAK and BD developed the full text screening form. LAK developed the data abstraction form. LAK, BD, YC, LCL, AA, LL, RM, SK, RW, JWB, and AD abstracted the data. LAK and AMK analysed the data. LAK, LH, RJPMS, and EAA interpreted the results. LAK, AMK, LH, RJPMS, and EAA drafted the manuscript. All authors revised and approved the final manuscript. EAA is the lead author and guarantor. LAK attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This study was funded by the Cochrane Methods Innovation Fund. The funder had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Competing interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: support from the Cochrane Methods Innovation Fund for the submitted work, no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: Data are available on reasonable request from the corresponding author at ea32@aub.edu.lb. Proposals requesting data access will need to specify how it is planned to use the data.

The lead author EAA (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: Lay information on the key results of the study will be made available on request from the corresponding author.

This manuscript has been deposited as a preprint.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academic Press, 2010.
- 2 Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials

(LOST-IT): systematic review. *BMJ* 2012;344:e2809. doi:10.1136/bmj.e2809

- 3 Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol* 2014;14:118. doi:10.1186/1471-2288-14-118
- 4 Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res* 2016;25:2977-83. doi:10.1007/s11136-016-1411-6
- 5 Hussain JA, Bland M, Langan D, Johnson MJ, Currow DC, White IR. Quality of missing data reporting and handling in palliative care trials demonstrates that further development of the CONSORT statement is required: a systematic review. *J Clin Epidemiol* 2017;88:81-91. doi:10.1016/j.jclinepi.2017.05.009
- 6 Powney M, Williamson P, Kirkham J, Kolamunnage-Dona R. A review of the handling of missing longitudinal outcome data in clinical trials. *Trials* 2014;15:237. doi:10.1186/1745-6215-15-237
- 7 Ibrahim F, Tom BD, Scott DL, Prevost AT. A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes. *Trials* 2016;17:272. doi:10.1186/s13063-016-1402-5
- 8 Kahale LA, Diab B, Khamis AM, et al. Potentially missing data are considerably more frequent than definitely missing data: a methodological survey of 638 randomized controlled trials. *J Clin Epidemiol* 2019;106:18-31.
- 9 Hallgren KA, Witkiewitz K, Kranzler HR, et al. In conjunction with the Alcohol Clinical Trials Initiative (ACTIVE) Workgroup. Missing data in alcohol clinical trials with binary outcomes. *Alcohol Clin Exp Res* 2016;40:1548-57. doi:10.1111/acer.13106
- 10 Guyatt GH, Ebrahim S, Alonso-Coello P, et al. GRADE guidelines 17: assessing the risk of bias associated with missing participant outcome data in a body of evidence. *J Clin Epidemiol* 2017;87:14-22. doi:10.1016/j.jclinepi.2017.05.005
- 11 Spinelli LM. An empirical comparison of Bayesian modelling strategies for missing binary outcome data in network meta-analysis. *BMC Med Res Methodol* 2019;19:86. doi:10.1186/s12874-019-0731-y
- 12 Akl EA, Kahale LA, Ebrahim S, Alonso-Coello P, Schünemann HJ, Guyatt GH. Three challenges described for identifying participants with missing data in trials reports, and potential solutions suggested to systematic reviewers. *J Clin Epidemiol* 2016;76:147-54. doi:10.1016/j.jclinepi.2016.02.022
- 13 Kahale LA, Guyatt GH, Agoritsas T, et al. A guidance was developed to identify participants with missing outcome data in randomized controlled trials. *J Clin Epidemiol* 2019;115:55-63. doi:10.1016/j.jclinepi.2019.07.003
- 14 Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. www.handbook.cochrane.org. 2011.
- 15 Akl EA, Kahale LA, Agarwal A, et al. Impact of missing participant data for dichotomous outcomes on pooled effect estimates in systematic reviews: a protocol for a methodological study. *Syst Rev* 2014;3:137. doi:10.1186/2046-4053-3-137
- 16 Kahale LADB, Diab B, Brignardello-Petersen R, et al. Systematic reviews do not adequately report or address missing outcome data in their analyses: a methodological survey. *J Clin Epidemiol* 2018;99:14-23. doi:10.1016/j.jclinepi.2018.02.016
- 17 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-81. doi:10.1016/j.jbi.2008.08.010
- 18 SPSS. *IBM SPSS Statistics for Windows. Version 21.0*. IBM Corp, 2013.
- 19 StataCorp. *Stata Statistical Software: Release 12*. In: Station C, ed. StataCorp LP, 2011.
- 20 Akl EA, Kahale LA, Agoritsas T, et al. Handling trial participants with missing outcome data when conducting a meta-analysis: a systematic survey of proposed approaches. *Syst Rev* 2015;4:98. doi:10.1186/s13643-015-0083-6
- 21 White IR, Welton NJ, Wood AM, Ades AE, Higgins JP. Allowing for uncertainty due to missing data in meta-analysis—part 2: hierarchical models. *Stat Med* 2008;27:728-45. doi:10.1002/sim.3007
- 22 White IR, Higgins JP, Wood AM. Allowing for uncertainty due to missing data in meta-analysis—part 1: two-stage methods. *Stat Med* 2008;27:711-27. doi:10.1002/sim.3008
- 23 Geng EH, Bangsberg DR, Musinguzi N, et al. Understanding reasons for and outcomes of patients lost to follow-up in antiretroviral therapy programs in Africa through a sampling-based approach. *J Acquir Immune Defic Syndr* 2010;53:405-11. doi:10.1097/QAI.0b013e3181b843f0
- 24 Geng EH, Emenyonu N, Bwana MB, Glidden DV, Martin JN. Sampling-based approach to determining outcomes of patients lost to follow-up in antiretroviral therapy scale-up programs in Africa. *JAMA* 2008;300:506-7. doi:10.1001/jama.300.5.506
- 25 White IR, Higgins JP. Meta-analysis with missing data. *Stata J* 2009;9:57-69. doi:10.1177/1536867X0900900104

- 26 Chaimani A, Mavridis D, Higgins JPT, Salanti G, White IR. Allowing for informative missingness in aggregate data meta-analysis with continuous or binary outcomes: Extensions to metamiss. *Stata J* 2018;18:716-40. doi:10.1177/1536867X1801800310
- 27 Higgins JP, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clin Trials* 2008;5:225-39. doi:10.1177/1740774508091600
- 28 Wasserstein RL, Schirm AL, Lazar NA. *Moving to a World Beyond "p < 0.05"*. Taylor & Francis, 2019.
- 29 Ebrahim S, Akl EA, Mustafa RA, et al. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol* 2013;66:1014-1021.e1.
- 30 Ebrahim S, Johnston BC, Akl EA, et al. Addressing continuous data measured with different instruments for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol* 2014;67:560-70. doi:10.1016/j.jclinepi.2013.11.014
- 31 Sullivan TR, Yelland LN, Lee KJ, Ryan P, Salter AB. Treatment of missing data in follow-up studies of randomised controlled trials: A systematic review of the literature. *Clin Trials* 2017;14:387-95. doi:10.1177/1740774517703319
- 32 Babic A, Tokalic R, Amílcar Silva Cunha J, et al. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. *BMC Med Res Methodol* 2019;19:76. doi:10.1186/s12874-019-0717-9
- 33 Shivasabesan G, Mitra B, O'Reilly GM. Missing data in trauma registries: A systematic review. *Injury* 2018;49:1641-7. doi:10.1016/j.injury.2018.03.035
- 34 Mavridis D, White IR. Dealing with missing outcome data in meta-analysis. *Res Synth Methods* 2020;11:2-13.
- 35 Hollestein LM, Carpenter JR. Missing data in clinical research: an integrated approach. *Br J Dermatol* 2017;177:1463-5. doi:10.1111/bjd.16010
- 36 Little RJ, Cohen ML, Dickersin K, et al. The design and conduct of clinical trials to limit missing data. *Stat Med* 2012;31:3433-43. doi:10.1002/sim.5519
- 37 Akl EA, Shawwa K, Kahale LA, et al. Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide. *BMJ Open* 2015;5:e008431. doi:10.1136/bmjopen-2015-008431

Supplementary information: additional tables, statistical notes, and results