Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide

Elie A Akl,1,2,3 Khaled Shawwa,1 Lara A Kahale,1,2 Thomas Agoritsas,3 Romina Brignardello-Petersen,3,4,5 Jason W Busse,3,6,7 Alonso Carrasco-Labra,3,4 Shanal Ebrahim,3,6,8,9 Bradley C Johnston,5,8 Ignacio Neumann,3,10 Ivan Sola,11 Xin Sun,3,12 Per Vandvik,13,14 Yuqing Zhang,3 Pablo Alonso-Coello,11 Gordon H Guyatt3,15

ABSTRACT

Objectives: We conducted a systematic survey of the methodological literature to identify recommended approaches for how and what randomised clinical trial (RCT) authors should report on missing participant data and, on the basis of these approaches, to propose guidance for RCT authors.

Methods: We defined missing participant data (MPD) as missing outcome data for trial participants. We considered both categorical and continuous outcome data. We searched MEDLINE and the Cochrane Methodology Register for articles in which authors proposed approaches to reporting MPD from RCTs. We selected eligible articles independently and in duplicate and extracted data in duplicate. Using an iterative process of discussion and revisions, we used the findings to develop guidance.

Results: Of 10 501 unique citations identified, 13 articles reporting on 10 approaches proved eligible. The identified approaches recommend reporting the following aspects (from most to least frequently recommended): number of participants with MPD (n=10), reasons for MPD (n=7), methods used to handle MPD in the analysis (n=4), flow of participants (n=3), pattern of missingness (eg, whether at random) (n=3), differences in rates of MPD between trial arms (n=2), differences between participants with and without MPD (n=2), results of any sensitivity analyses (n=2), implication of MPD on interpreting the results (n=2) and methods used to prevent missing data (n=1). We propose a guide with nine items related to reporting the number, reasons, patterns, analytical methods and interpretation of MPD.

Conclusions: Most identified approaches invite trial authors to report the extent of MPD and the underlying reasons. Fewer approaches focus on reporting missingness patterns, methods for handling MPD and implications of MPD on results. Our proposed guidance could help RCT authors to better report, and readers to better identify participants with missing data.

Strengths and limitations of this study

- First systematic survey addressing recommendations for the reporting of missing participant data in randomised clinical trials.
- Explicit eligibility criteria with an appropriate search for relevant English language articles.
- Systematic approaches to study selection, data abstraction and data synthesis.
- A limitation in excluding non-English studies.
- We did not implement duplicate data extraction, but a second reviewer checked all the extracted data for accuracy.

BACKGROUND

Missing participant data is common in randomised clinical trials (RCT). A methodological survey of the top five general medical journals found that 191 of 235 (87%) of published trials reported missing participant data (MPD) for the primary outcome. The median percentage of participants with missing data was 6% (IQR 2–14%).1 Of the 191 trials reporting MPD, a third lost statistical significance when making plausible assumptions about the outcomes of missing participants.1

Systematic reviews, health technology assessments and clinical practice guidelines based on results from RCTs are vulnerable to bias that may result from MPD in the primary trials. In order to assess risk of bias resulting from MPD, consumers of the medical literature must identify the number and characteristics of trial participants for whom outcome data are missing. Reports of RCTs do not, however, always include this information in a consistent and clear manner. Indeed, Sylvestre et al2 found that

Correspondence to Dr Elie A Akl; ea32@aub.edu.lb


Received 19 April 2015
Revised 6 November 2015
Accepted 9 November 2015

For numbered affiliations see end of article.
information on missing values was not present in one-quarter of 93 Health Technology Assessments trial reports. Moreover, contact with authors of primary studies in the aforementioned survey revealed that unclear reporting was responsible for most inaccuracies in data abstraction.1

The Consolidated Standards of Reporting Trials (CONSORT) statement recommends standards for reporting of the findings of randomised trials.3 The standards address, among other issues, the reporting of loss to follow-up in trials. These ‘evidence-based’ recommendations were published in 2010, and would benefit from the identification on the current best available evidence on the topic.

The main objective of this study was to systematically review the methodological literature to identify recommended approaches for how and what RCT authors should report on missing participant data and, on the basis of these approaches, to propose guidance for RCT authors. This study was part of a larger project addressing the issue of missing participant data in trials and systematic reviews.

METHODS
Definition
Missing participant data refers to missing outcome data for trial participants. This does not include missing participant baseline characteristics (eg, patient age).

Eligibility criteria
We included articles that met the following criteria:
Inclusion criteria

▶ The paper discussing methods or conceptual approaches to addressing how and what RCTs should report on missing participant data. A typical example would be a paper on reporting standards such as the CONSORT statement.3 A paper describing challenges and solutions, or reviewing the literature for guidelines on how RCT should report on missing participant data would also be potentially eligible.
▶ The paper should have devoted at least two paragraphs to discuss the topic of interest (criterion applied when reviewing the full texts).
▶ The paper could have considered reporting of categorical and/or continuous data.

Exclusion criteria

▶ Reports of systematic reviews or of trials.
▶ Papers discussing how to prevent, minimise, handle, analyse or assess risk of bias associated with missing participant data.
▶ Papers written in languages other than English.

Search strategy
Given that the focus of the study was on reporting in health-related trials, as opposed to dealing with MPD in statistical analyses, our search focused on the medical literature as opposed to the statistical literature. In August 2014, we searched MEDLINE, from its inception date using the OVID interface. We also searched the Cochrane Methodology Register. A researcher with experience in developing literature search strategies (IS) developed an initial search strategy. We subsequently used relevant articles identified through the pilot search to refine the strategy (see online supplementary appendix 1). In order to be comprehensive, we reviewed the CONSORT statement with its extensions.3-6

Article selection
Using a web-based systematic review software (SRDistiller), reviewers (LAK, TA, RB-P, JWB, AC-L, SE, BCJ, IN, IS, XS, PV and YZ) conducted screening in pairs and independently: first they screened titles and abstracts, and we obtained the full texts for those judged as potentially eligible by at least one of the two reviewers. Then, they screened these full texts for eligibility, compared their judgements and resolved disagreements by discussion, or, if necessary, with the help of a third reviewer (EAA). In order to ensure clarity and consistency, and prior to initiating the article selection process, we conducted calibration exercises and pilot tested the screening forms on a number of potentially eligible articles.

We calculated agreement for full-text screening stage using the k statistic. We interpreted the degree of agreement between pairs of reviewers according to the criteria proposed by Landis and Koch7 (k values of 0–0.20 represent slight agreement; 0.21–0.40 fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement; and >0.80 values represent almost perfect agreement).

Data abstraction and presentation
One reviewer (KS) abstracted data from included articles. A second reviewer (EAA) verified all the abstracted results. We used an iterative process to optimise the presentation of the abstracted data. We abstracted data from one eligible article at a time into a table with columns listing categories of reporting recommendations. We started with a preliminary list of categories including: number of participants with MPD, reasons for MPD and participant flow diagram. With every additional article being abstracted, we modified those categories as needed to integrate all relevant information from that article. We followed this approach until we abstracted data from all eligible articles. We conducted this process through face-to-face meetings. The remaining authors provided suggestions on how to improve data presentation. We used these recommendations as the basis for developing a guide for trialists.

Developing the guide
The two reviewers who abstracted the data developed an initial draft guide based on the identified recommendations in a number of face-to-face meetings (average of 2–3 times/week over a 4-month period from start of data abstraction up to finalisation of the guide). They used
an iterative process of discussion and revisions to refine the draft. Specifically, they reviewed one eligible article at a time and modified the draft to integrate any new concepts in a coherent way. They followed this approach until they reviewed all eligible articles. The remaining members of the team reviewed and commented on the draft guide through email communication. These team members include clinical epidemiologists with extensive experience in clinical trials and systematic review methodologies. The discussion was informed by the team members’ previous work on dealing with missing participant data in published trials. One of the challenges that we encountered was the inconsistency of the terminology used across papers to refer to the same concepts. While the team had to agree on which terminology to use, we decided, for transparency and accuracy purposes, to report in an appendix the terminology used in each included paper.

RESULTS
Our search strategy identified 10,572 citations, of which 13 proved eligible (figure 1). Agreement between authors for study eligibility was almost perfect ($\kappa=0.95$).
The 13 articles described 10 approaches; 1 of the approaches was the CONSORT statement, and three articles reported CONSORT extensions. These extensions were for patient reported outcomes (PROs), harm and cluster trials.

**Recommended approaches**

We report in online supplementary appendix 2 the recommendations of each included paper. The text in the appendix reproduces the paper’s own terminology for referring to missing participant data. The recommendations can be summarised as follows:

- Report methods used to prevent MPD;
- Report number of participants with MPD;
- Report differences in rates of MPD between trial arms;
- Report the reasons for MPD;
- Report a flow of participants;
- Report any differences between participants with and without MPD;
- Report pattern of missingness (eg, whether at random);
- Report methods for handling MPD in analysis;
- Report results of any sensitivity analyses;
- Discuss implication of MPD on interpreting the results.

We report in online supplementary appendix 3 the definitions of the different patterns of missingness, as well as the terminology used by each paper to describe the different reasons for missing participant data. Papers used a range of terms and different approaches to classifying missing data. A number of papers used terms that describe the underlying cause of missingness:

- Health status related: for example, death, illness, progressive disease (n=4);
- Participant choice related: lack of interest, lack of time, bothered by question (n=2);
- Technically related: questionnaire not given, wrong questionnaire instructions, transportation problem (n=2).

A number of papers used terms that describe the pattern of missingness (n=5):

- Informativeness (non-random) censoring versus non-informative (random) censoring;
- Missing at random versus not missing at random versus missing completely at random;
- Intermittent or non-monotone missingness.

One paper used terms that describe who caused the missingness: researcher initiated (eg, removal of participants) versus participant related (eg, withdrawal).

**Table 1** describes each of the 10 approaches which specific recommendations are covered (only as frequency). Three articles specifically address issues in reporting missing data in trials using continuous outcome measures such as PROs. The remaining articles apply to either categorical or continuous outcome measures. The identified approaches recommend reporting the following aspects (from most to least frequently recommended): number of participants with MPD (n=10), reasons for MPD (n=7), methods used to handle MPD in the analysis (n=4), flow of participants (n=3), pattern of missingness (eg, whether at random) (n=3), differences in rates of MPD between trial arms (n=2), differences between participants with and without MPD (n=2), results of any sensitivity analyses (n=2), implication of MPD on interpreting the results (n=2), and methods used to prevent missing data (n=1).

**Proposed guide**

**Box 1** presents our proposed guide on how RCT authors should report missing participant data. These include items relevant to the report of both categorical and continuous variables as well as items specific to the report of continuous variables. The guide does not specify the format of reporting, which could be narrative, tabular, or graphical (eg, study flow).

**DISCUSSION**

The majority of approaches to reporting missing data recommend that trial authors report the extent of missing participant data and the underlying reasons. Fewer approaches focus on patterns of missingness, methods for handling MPD and implications of MPD on results.

This guidance builds on, and complements the CONSORT statement, as it relates to MPD. CONSORT wisely recommends reporting a flow diagram of the progress of participants through the phases of the trial by study group, including loss to follow-up with reasons, and the number of participants excluded from the analysis. Our proposed guidance is more specific (eg, addressing missing data for each outcome separately) and wider in scope (eg, handling MPD in the main analysis and in any sensitivity analysis, evaluating impact of MPD on interpretation of results). Publication or sharing of trial raw individual participant data, would automatically allow meeting many of the recommendations (eg, participants with missing data by arm, by outcome, or by item; baseline characteristics of participants with missing data).

The recently published SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) statement provides recommendations for a “minimum set of scientific, ethical, and administrative elements that should be addressed in a clinical trial protocol.” Although not strictly eligible for this study, the statement highlights the importance in explicit reporting of MPD, starting with the protocol. For example, it invites trialists to prespecify the methods of statistical analysis of the primary outcome and how missing data will be handled. This includes details of the planned methods for imputing MPD, including which variables will be used in the imputation process. The guidance also includes outlining the planned approach to making the final methodological choices when these cannot be prespecified (eg,
### Table 1: Summary of proposed approaches for reporting MPD

<table>
<thead>
<tr>
<th>Author</th>
<th>Methods used to prevent missing data</th>
<th>Number of participants with MPD</th>
<th>Differences in rates of MPD between trial arms</th>
<th>Reasons for MPD</th>
<th>Flow of participants</th>
<th>Differences between participants with and without MPD</th>
<th>Pattern of missingness (eg, whether at random)</th>
<th>Methods for handling MPD in analysis</th>
<th>Results of any sensitivity analyses</th>
<th>Implication of MPD on interpreting the results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staquet et al&lt;sup&gt;13&lt;/sup&gt;</td>
<td>✓✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bernhard et al&lt;sup&gt;12&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Troxel et al&lt;sup&gt;20&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Liu et al&lt;sup&gt;11&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Amico et al&lt;sup&gt;21&lt;/sup&gt;</td>
<td>✓**</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sterne et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Polit and Gillespie&lt;sup&gt;24&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Al-Shurafa et al&lt;sup&gt;25&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gewandter et al&lt;sup&gt;26&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CONSORT‡&lt;sup&gt;3&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Number of studies recommending it</td>
<td>1</td>
<td>10</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

*These approaches further recommended reporting missing data by study arm.
†Troxel et al recommended examining patient-related and institution-related factors affecting missing data rates descriptively or by using logistic regression models.
‡Elements recommended by CONSORT extensions, not already included in the main statement: (1) extension for harms, none; (2) extension for cluster trials: reporting for each group, the missing participant data for both clusters and individual cluster members; (3) extension for patient reported outcomes, reporting outcome data at baseline and at subsequent time points, interpreting any supportive (eg, sensitivity) analyses.

MPD, missing participant data.
the method of handling missing data which might depend on examining patterns of ‘missingness’ when data become available).

While the focus of this paper is to improve the reporting of MPD to assist in their handling in systematic reviews, avoiding or minimising MPD remains the ideal solution for MPD.\(^\text{15–17}\) This shifts the burden of addressing the problem from statisticians to trialists. There has been a number of prominent guidance on this by a number of bodies such as the Food and Drug Administration.\(^\text{18}\)

**Strengths and limitations**

To the best of our knowledge, this is the first systematic survey addressing recommendations for the reporting of MPD in RCTs. Strengths of this survey include explicit eligibility criteria, an appropriate search for relevant English language articles, and systematic approaches to study selection, data abstraction and data synthesis. One limitation of the review is the exclusion of non-English studies. Although there is evidence that exclusion of non-English studies might result in the loss of an appreciable number of eligible studies in clinical systematic reviews,\(^\text{19}\) this may be less of an issue for methodological reviews. We did not implement duplicate data extraction, but a second reviewer checked all extracted data for accuracy. Also, we did not keep track of the frequency of agreements and disagreements regarding which items are included in the final version of the guide.

**Conclusion**

We have summarised the recommended approaches for how trialists should report MPD, and proposed guidance based on our findings. Our findings have implications for trialists as well as editors of medical journals. Both of these groups may wish to consider adhering to this guidance when reporting trials to help the users of the medical literature to adequately identify participants with missing data to judge the validity of trial findings. Adherence to our suggestions would also allow systematic reviewers to identify MPD in order to conduct meta-analyses that adequately take them into account. The authors of the CONSORT statement may consider integrating our guidance in a future update of that statement.

Our findings have implications also for future research. There is a need to assess to what extent reports of RCTs adhere to those reporting recommendations, particularly to assess response to any initiatives to improve MPD reporting. More generally, there is a need for more research on how to prevent, minimise, handle, analyse and assess risk of bias associated with MPD.

**Author affiliations**

1. Department of Internal Medicine, American University of Beirut, Beirut, Lebanon
2. Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon
3. Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Ontario, Canada
4. Evidence Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, Santiago, Chile
5. Institute of Health Policy, Management & Evaluation, University of Toronto, Toronto, Canada
6. Department of Anesthesia, McMaster University, Hamilton, Canada
7. The Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, Canada
8. Department of Anesthesia & Pain Medicine, the Hospital for Sick Children, Toronto, Canada
9. Department of Medicine, Stanford Prevention Research Center, Stanford University, Palo Alto, USA
10. Department of Internal Medicine, School of Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile
11. Iberoamerican Cochrance Centre, Biomedical Research Institute Sant Pau-CIBER of Epidemiology and Public Health (CIBERESP- IIB Sant Pau), Barcelona, Spain
12. Clinical Research and Evaluation Unit, Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu, China
13. Norwegian Knowledge Centre for the Health Services, Oslo, Norway
Contributors EAA, PA-C and GHG contributed to the conception and design. EAA and IS were responsible for design of search strategy. LAK, TA, RB-P, JWB, AC-L, SE, BCJ, IN, IS, XS, PV and YZ selected the paper. EAA and KS contributed to data abstraction, data synthesis and manuscript drafting. EAA, KS, LAK, TA, RB-P, JWB, AC-L, SE, BCJ, IN, IS, XS, PV, PA-C and GHG were responsible for interpretation of results. EAA, KS, LAK, TA, RB-P, JWB, AC-L, SE, BCJ, IN, IS, XS, PV, YZ, PA-C and GHG were responsible for manuscript review and approval.

Funding This paper is part of a project on addressing missing trial participant data in systematic reviews funded by the Cochrane Collaboration.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access 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/

REFERENCES


Appendix 1: Search Strategy

Search Strategy- Medline

#1 "drop-out*".m_titl.

#2 missing.m_titl.

#3 "withdraw*".m_titl.

#4 (los* and follow*).m_titl.

#5 per protocol.m_titl.

#6 intention-to-treat.m_titl.

#7 intent-to-treat.m_titl.

#8 ITT.m_titl.

#9 (exclusion or exclusions).m_titl.

#10 excluded.m_titl.

#11 or/1-11

#12 ((randomized controlled trial or controlled clinical trial).pt. or randomized.ab. or placebo.ab. or drug therapy.fs. or randomly.ab. or trial.ab. or groups.ab.) not (exp animals/ not humans.sh.)

#13 11 and 12
Search Strategy – Cochrane Library

attrition or (drop out) or missing or withdraw* or (loss* and follow*) or (per protocol) or (intention to treat) or (intent to treat) or ITT or (exclusion or exclusions or exclude):ti,ab,kw

Restricted to: Methods Studies and Technology Assessments
Appendix 2: The recommendations of each included paper. The text here reproduces the paper’s own terminology for referring missing participant data terminology

Staquet et al. propose general reporting guidelines for clinical trials that report a quality of life measurement:[13]

- Clearly state the methods by which missing data were defined and analyzed;
- Report the number of patients with missing data separately in each arm of the trial;
- Document separately the types of missing data by cause, including: non-completion of the questionnaire due the death of the participant; non-completion of the questionnaire for reasons other than death; and non-response to items in the questionnaire;
- Compare and comment on the percentages of missing data by item in the questionnaire and by treatment group, with focus on the major endpoints as specified in the protocol.

Bernhard et al. discuss issues pertaining to missing quality of life data in cancer clinical trials.[12] They provide an estimation of the magnitude of the problem, and approaches to its prevention and solution. They state that trialists “have an obligation to provide enough information to allow the reader to assess the quality of the study”. They also suggest that good trial reports should allow the following three questions to be answered:

- How many missing data are there?
- Why were the data missing?
How will the missing data affect the interpretation of the study results?

In addition, the authors recommend presenting separately the overall form submission rates for each scheduled assessment, with the aim of evaluating trends over time. They also recommend evaluating the quality of the completed questionnaire in terms of the item response rates.

Troxel et al. discuss statistical analysis of quality of life with missing data in cancer clinical trials. [20] They outline how missing data are often described as either 'dropout' or 'intermittent'. The authors make the following suggestions on reporting missing data:

- Report detailed summaries of the completeness of study data with focus on the amount and reasons for missing data at each scheduled assessment, in order to allow for the evaluation of trends over time;
- Compare treatment group differences in quality of life compliance when applicable;
- Examine patient- and institution-related factors affecting missing data rates descriptively or by using logistic regression models;
- Present the data separately for patients with different reasons for dropout

Present the result of studies graphically, comparing scores for patients whose self-assessment was missing with those for whom it was not

Liu et al. review the International Conference on Harmonization (ICH) guidelines and the
Committee for Proprietary Medicinal Products (CPMP) provide guidance regarding reporting missing data in clinical trials [21]. The authors recommend reporting the extent, reasons, and patterns of missing data. They also recommend conducting and reporting the results from additional sensitivity analyses if the extent and pattern of missing data deviate substantially from those anticipated.

Amico et al. discuss attrition in research on antiretroviral therapy.[22] They also review recommendations on attrition pertaining to scientific rigor in longitudinal intervention trials, including the following:

- Clearly report the flow of participants
- Quantify and assess attrition for potential differential rates between the different treatment arms
- Categorize reasons for attrition
- Make transparent the assumptions for missing data and strategies to replace them

Sterne et al. review why missing data may lead to bias and information loss in clinical and epidemiological research.[23] They discuss the use of multiple imputations, its potential pitfalls, and guidance for conduct and reporting. They propose the following recommendations:

- Report the number of missing values for each variable of interest
- Provide reasons for missing values when possible
• Indicate the number of participants excluded because of missing data when reporting the flow of participants
• Describe reasons for missing data in terms of other variables
• Clarify whether there are important differences between individuals with complete and incomplete data
• Describe the types of analysis used to account for missing data and the assumptions made

The authors provide additional reporting guidelines for analyses based on multiple imputations (see box 2 in the original paper for further details).

Polit et al. discuss the definitions of intention to treat (ITT), and recommend strategies for implementing ITT in clinical trials [24]. They also discuss types of missing outcome data and offer the following reporting suggestions:

• Report any experimental-control group differences in attrition
• Description of characteristics distinguishing study completers from dropouts
• Discussion of the likely pattern of missingness (i.e., missing completely at random (MCAR) vs. missing at random (MAR) vs. missing not at random (MNAR))
• Report the results of any sensitivity analyses
• Discussion of the possible implications of missing values on estimates of intervention effects (e.g., high rates of missingness, or pattern of missingness that may not be at random or completely at random).
AlShurafa et al. conducted a systematic review of how authors of methodology articles define ITT when outcome data are not available for all participants and how they recommend handling missing outcome data when conducting an ITT analysis.[25] The authors provide “essential components to report in RCTs with respect to the analysis” that include items related to reporting the handling of missing outcome data:

- Whether there are individuals with missing outcome data;
- Whether those individuals are not considered in the analysis (complete/available case analysis);
- Whether the outcomes of those individuals were imputed; and in that case the imputation(s) used.

Gewandter et al conducted a systematic review to assess the frequency with which RCTs published in three pain journals reported strategies to prevent missing data, the number of completers, and statistical methods to handle missing data [26]. The authors made the following recommendations:

- Report methods used to prevent missing data in clinical trials
- Report the numbers of participants who were randomized, who completed the trial, and who were included in the analyses
- Explicitly state the number of participants who provided complete data in trials in which
the primary outcome variable is defined using data from multiple time points

- Describe the distribution of the percentage of data missing.

The CONSORT 2010 statement provides authors with a standardized approach for reporting trial findings. [3] The statement recommends reporting a flow diagram of the progress of participants through the phases of the trial by study group, including loss to follow-up with reasons, and the number of participants excluded from the analysis.

CONSORT-Patient Reported Outcomes (PRO) extension[6] provides the following additional recommendations:

- State the statistical approaches for dealing with missing data.
- Make the number of PRO outcome data at baseline and at subsequent time points transparent
- Include information on the reason for missing PRO forms (e.g., lack of questionnaire return, unavailable translations of the questionnaire, or other known reasons).
- Provide information regarding reasons for missing PRO forms in either a tabulated form (by treatment group), or in footnotes of the CONSORT flow diagram.
- Discuss the potential reasons for missing PRO data in relation to the clinical context and implications for interpretation of findings.
- Discuss the interpretation of any supportive (e.g., sensitivity) analyses undertaken.
The CONSORT harm extension provides guidance for authors of trials: includes a set of recommendations for the proper reporting of harms in RCTs [5]. The extension provided the following recommendations specific to reporting of MPD:

- Report participants who are non-adherent or lost to follow-up as this may reflect intolerance of the intervention.
- Report the denominator for each analysis (that is, which participants and what follow-up time count toward total exposure to the allocated treatment) in RCTs in which time-on-treatment differs from total follow-up.

The CONSORT extension for cluster trials recommends reporting for each group, losses and exclusions for both clusters and individual cluster members [4].
**Appendix 3:** Definition of the types of missingness and terminology used by each paper to describe the different reasons for missing participant data, with specific terms underlined.

<table>
<thead>
<tr>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing completely at random: missingness of an outcome is not related to any observed or unobserved variables; the missingness probability is independent of all previous, current and future assessments.</td>
</tr>
<tr>
<td>Missing at random: missingness of an outcome may be related to observed or unobserved variables, but is not related to the actual value of the outcome, conditional on the observed variables; the missingness probability does not depend on the missing values.</td>
</tr>
<tr>
<td>Informatively missing: missingness of an outcome is related to the value of that outcome, even conditional on other observed variables; the missingness probability may depend on the unobserved values.</td>
</tr>
</tbody>
</table>

**References:**


Terminology used to describe the different reasons for missing participant data

Staquet 1996 [13]

- Missing data due to non-completion of the questionnaire resulting from the death of the patient;
- Missing data due to non-completion of the questionnaire for reasons other than death
- Missing data due to non-response to items

It should be specified what missing data are due to informative (non-random) censoring (i.e. missing data due to the patient’s health state or to his particular treatment) or to non-informative (random) censoring.

Bernhard 1998 [12]

In some circumstances, data are unavoidably missing, as for example when patients have died or are too ill to complete forms.

Obviously, certain sources of missing QOL data, such as attrition due to death and withdrawal of some patients from QOL assessment due to progressive disease or treatment-related toxicity, are unavoidable.

A number of other factors can account for missing data: questionnaires are inadvertently not
given to the patient, the wrong questionnaires are given, or questionnaires are administered at the wrong time; the patient is either not given instructions or not appropriately instructed in how to complete the questionnaire; questionnaires are not routinely inspected for missing responses after collection.

Deteriorating health status is one of the patient factors that both generates missing QOL data and is not very amenable to intervention.

In case the patient refused to complete the questionnaire, give reason:

- Felt too ill
- Lack of interest or time
- Complained of burden
- Bothered by questions
- Other (specify):

**Troxel 1998 [20]**

QOL assessments are often not obtained because of negative events experienced by patients such as treatment toxicities, disease progression, or even death.

Missing data are often described as either ‘dropout’ or ‘intermittent’. Dropout occurs when a subject, once missing an assessment, is never observed again. Intermittent or non-monotone missingness occurs when a subject misses an assessment but is later observed.
Paper discusses: missing at random (MAR), missing not at random (MNAR), and missing completely at random (MCAR)

Liu 2006 [21]

A selection model is a natural choice, where patients withdrew from the study because of worsening in treatment responses in FPG and HbA1C

In the pattern mixture model, parameter $f$, the parameter conditional on the missing data pattern, is the primary concern, i.e., probability distribution of the response depends on dropout status and different response patterns can be modeled for patients who drop out or continue for different reasons

Discussed MCAR, MAR and NMAR

- Examples of MCAR include patients who moved away due to relocation (not reasons related to study outcomes), closing of the study and late entry of patients administratively ‘censored’
- Examples of MAR include the predefined dropout process based on recorded biomarker values. For example, in a diabetes study, the protocol pre-defined a withdrawal criterion as when a patient’s recorded fasting plasma glucose (FPG) level exceeds 270 mg/dl.
- Dropouts due to drug-related adverse events (AEs) such as weight gain, GI intolerance or hypoglycemic events could also be considered as MAR. An example of NMAR is missing FEV1 in asthma trials when patients withdraw due to acute worsening
conditions, and the reason for the missing FEV1 measurements are likely related to the unobserved FEV1 values, and hence not missing at random.

Patients are likely to withdraw from the treatment due to lack of efficacy.

Missing data/dropouts

In diabetes clinical trials, it is usually anticipated that patients would withdraw if their glycemic conditions worsen because they are not on any anti-hyperglycemic drug (in placebo-controlled trials) or they develop resistance or do not respond to a certain class of anti-hyperglycemic drugs.

The second type of missing data includes those withdrawals due to treatment-related AEs that are likely to be MAR provided that treatment is included in the analysis model. The third type of missing data includes those due to administrative reasons, such as patients’ moving away, which would be MCAR. Other types of missing data are caused by patients withdrawing consent (refuse to be on Statin), protocol violations (e.g., did not use double barrier), etc., which could be MAR or NMAR.

Amico 2008 [22]

Further classifications for attrition in terms of reasons (eg, requested withdrawal vs lost) or other categorization systems were used in only 18 (35%) of the studies, and partially used in 10 (20%).
Sterne 2009 [23]

Reasons for missing data are commonly classified as: MAR, MCAR and NMAR.

Polit 2010 [24]

Yet, the advantages of an experimental design can be undermined by removals, withdrawals, and subject losses, which can nullify the initial equivalence of the experimental and control group.

Two categories:

1- Researcher-initiated: removal of subjects who were not really eligible, non-compliers/dropouts/protocol violators. Removal of subjects who actually received the alternative treatment (crossovers)

2- Subjects-related: death, severe illness/disability, severe pain, lack of interest or time, preference for other treatment, transportation problems, Unable to locate subject to obtain post-randomization outcome data.

Discuss the mechanism of missingness (MAR, MCAR and MNAR)

- For example, an accidental death would typically result in outcomes that are MCAR

If the intervention condition is especially attractive, however, there is a risk that control group
participants will drop out of the study because of their disappointment at not receiving it, or they may seek similar alternative treatments.

In many studies, outcome data are missing for people who drop out of the study because of time constraints or other barriers.

Alshurafa 2012 [25]

Missing outcome data (MOD) mainly due to patients lost to follow-up for the primary analysis.