Marie Curie

Care and support through terminal illness

Missing data in palliative and end of life care trials

Guidance on how to reduce, handle and report incomplete data

June 2022

FUNDED BY NIHR | National Institute for Health Research

HULL YORK MEDICAL SCHOOL

Prepared on behalf of Marie Curie Research by:

Dr Jamilla Hussain, Wolfson Palliative Care Research Centre, Hull York Medical School, University of Hull

Professor Tim J Peters, Bristol Medical School and Bristol Dental School, University of Bristol and Chair of the Marie Curie Research Funding Committee

Professor Miriam Johnson, Wolfson Palliative Care Research Centre, University of Hull

Professor Anthony Byrne, Marie Curie Palliative Care Research Centre, Cardiff University

Professor Nancy Preston, International Observatory on End of Life Care, Lancaster University

Professor Andrew Haines, Department of Public Health, Environments and Society and Department of Population Health, London School of Hygiene & Tropical Medicine

Dr Kathy Seddon, Marie Curie Research Voice and Research Partner at the Wales Cancer Research Centre

Professor Ian R White, MRC Clinical Trials Unit at UCL, University College London

Contents

Acknowledgements

The writing group would like to thank all participants of the 2017 workshop for helping to develop the guidance.

Marie Curie provided financial and administrative support for the workshop.

Dr Jamilla Hussain was supported by an NIHR Doctoral Research Fellowship grant and NIHR Clinical Lectureship.

Prof Ian White was supported by the Medical Research Council Programme MC_UU_00004/07.

Our particular thanks go to all members of the Marie Curie Research Strategic Advisory Committee (see Appendix 1), chaired by Professor Sir Andy Haines between 2014 and 2019, for their input in conceiving and shaping the workshop and advising on the outcomes and dissemination strategy.

Glossary

Auxiliary variables – variables not in the main statistical model, but which are associated with missing data. They can be used to strengthen the plausibility of a missing at random assumption.

Bias – in statistics, bias refers to a systematic error or deviation from the truth.

Data truncated due to death – when data are not collected because the participant has died. Compared with missing data in those alive, missing data due to death are not considered to be "missing" as such, but rather "undefined", because how or whether to define the value once the participant has died needs to be considered.

Estimand – a description of what needs to be estimated to address a specific research question. An estimand should include five attributes: the treatment conditions being compared; the participant group of interest for the research question; the outcome variable of interest; how intercurrent events, such as the participant stopping the treatment, will be addressed; and the population-level summary that provides a basis for comparison between treatment conditions[1].

Estimate – numerical estimate of the estimand using a specified method of analysis (estimator) and the data that were observed[1].

Explanatory variables – variables that are used to predict or explain differences in the outcome variable. In a trial, this would include, for example, whether or not the participant received the new treatment being investigated, and often participant attributes collected at the start of the study (**baseline characteristics**).

External validity – refers to the extent to which the results can be applied or generalised to individuals other than those included in the study. This is sometimes referred to as the generalisability of the findings.

Feasibility studies – studies done before the main trial in order to determine whether the trial can be undertaken successfully. They are used to estimate important parameters that are needed to design the main study, such as recruitment and retention rates[2].

Internal validity – refers to the extent to which the observed results represent the truth in the population being studied, i.e. how well the study has prevented bias.

Main trial – the main clinical trial aiming to assess the effectiveness of the intervention of interest. In drug development, these are also known as phase 3 randomised trials.

Mechanism of missing data – method by which the data came to be missing. Three mechanisms are commonly referred to:

- **Missing completely at random (MCAR)** – when missingness is nothing to do with the participant, e.g. a blood sample is dropped in the laboratory
- **Missing at random (MAR)** – when missingness is related to the participant and can be predicted from other information about them
- **Missing not at random (MNAR)** – when missingness is specifically related to the data that are missing[3].

Missing data – information that would be meaningful to help answer the research question that was intended to be collected, but for whatever reason was not. This includes data that are missing because the participant could no longer provide any data, because, for instance, they were too unwell (known as **unit-level missing data**). In addition, there can be partially missing data, for example when the participant answers some but not all of the questions in a questionnaire (known as **item-level missing data**).

Missing data sensitivity analyses – an assessment of the sensitivity of the findings in respect of different assumptions about the missing data mechanism.

Outcome variables – variables that typically describe an individual's health or quality of life, which may be affected by treatment. Often trials have a small number of primary outcome variables that are the main outcomes researchers are interested in assessing and that are held to potentially influence policy and/or practice. A potentially larger number of secondary outcome variables are also assessed, which are not as important as the primary outcome but are still of interest when evaluating the effect of an intervention.

Palliative care – treatment, care and support for individuals with a progressive life-limiting illness with no possibility of remission, receiving holistic inter-disciplinary treatment focusing on quality of life[4]. This includes, but is not limited to, end of life care.

Performance status – a score that assesses an individual's general wellbeing and ability to undertake activities of daily living. Higher scores reflect better functional ability.

Pilot study – a miniature version of the main trial performed to determine whether the components of the main study can work together[2]. Pilot studies can be **internal** (if the data are intended to form part of the trial itself) or **external** (when the plan is that they will not form part of the final data analysis).

Power of a study – a study's ability to detect a difference between treatment groups, if one exists.

Protocol – the written plan to be followed in a study. For a clinical trial, this will include, among other things, aims and objectives, design, general methodology and statistical considerations.

Proxy – an individual nominated by the participant to provide consent and/or data on their behalf if they are unable to do so. This is usually a family member, carer or healthcare professional.

Statistical model – a summary of the relationship between the explanatory and outcome variables in a dataset.

Variables – attributes that describe an entity, such as a participant in a clinical trial, which can vary from one entity to another. Examples include age, gender and level of pain.

Executive summary

Missing data are observations (i.e. information) that would be meaningful to help answer a research question and which were intended to be collected, but for whatever reason were not.

Large amounts of missing data are found in palliative and end of life care studies. These can reduce the ability of a study to detect whether a new treatment is helpful or not, and how applicable the findings are to different types of people. Crucially, missing data can also affect how truthful the findings of a study are, as they can introduce bias.

Who is this guidance for?

This document provides guidance on how to reduce, handle and report missing data in palliative and end of life care trials. It aims to inform interested patients and carers, patient and public involvement (PPI) partners, clinical teams, researchers, funders and policymakers about how missing data should be addressed throughout the course of a study and how to evaluate the risks missing data pose to research findings.

Although this guidance focuses on palliative care studies, many of the recommendations will be relevant to other areas of healthcare research.

How were the guidelines developed?

The guidelines are based on research evidence, other guidelines and practice both within and outside of palliative and end of life care research. This information was synthesised and shared with participants at a missing data workshop hosted by Marie Curie. Participants included PPI partners, clinicians, researchers, statisticians and methodologists, who helped refine and develop the guidance presented[5].

Summary of the guidance

The guidance is structured as follows:

- Part A: how to **reduce** missing data in palliative and end of life care trials
- Part B: how to **handle** missing data in palliative and end of life care trials
- Part C: how to **report** missing data in palliative and end of life care trials.

A. How to reduce missing data in palliative and end of life care trials

All statistical methods to analyse datasets with missing data have limitations – therefore, reducing missing data is essential. This is because statistical techniques are often based on assumptions that cannot be verified, as the true values of the data that are missing are not known. The box below summarises the recommendations for reducing missing data in palliative and end of life care trials.

A summary of recommendations for reducing missing data in palliative and end of life care trials

1) **Prepare and plan for how to reduce missing data at the trial design and protocol development stage.** This includes developing a flexible and inclusive study design, consulting members of the multidisciplinary team involved in conducting a trial on how to reduce missing data, reducing the trial burden and evaluating strategies to reduce missing data.

2) **Resource the trial adequately to minimise missing data.**
This includes funding for data collection across settings and the use of different modalities of data collection, incentives for sites to provide complete data and reasons for missing data, and recruitment of staff with a good track record for data collection.

3) **Train all research staff to understand the risks posed by missing data and how to minimise missing data.**

4) **Discuss the value of complete data and how to reduce missing data with participants before they consent to enter the trial.**
This includes exploring their concerns about the data collection process and informing them why each outcome is being collected, the importance of complete data and why collecting the reasons for missing data is important. Also gain consent for the use of proxies and/or access to their medical records if they are unable to provide data.

5) **Collect the reasons for missing data.**

6) **Distinguish participants who want to withdraw from providing any further data from participants who wish to withdraw from part of the study protocol but consent to ongoing data collection or access.**

7) **Monitor and address missing data during the trial.**

B. How to handle missing data in palliative and end of life care trials

Even with careful consideration of how to reduce missing data, some data will be missing in a large proportion of studies. It is crucial that such data are handled with a principled statistical approach that reduces bias as much as possible. The box below summarises the recommendations for handling missing data in palliative and end of life care trials.

A summary of recommendations for handling missing data in palliative and end of life care trials

1) Include a statistician in the trial team during the design, conduct and analysis stages of the study.

2) Decide how missing data will be handled in the design and conduct of the study and in its analysis, and report these decisions in the protocol and statistical analysis plan.

3) Prepare for missing data analyses at the trial design stage.
This includes collecting the reasons for missing data and considering whether any auxiliary variables should be collected.

4) Inflate the sample size to account for expected missing data in order to achieve the number of participants necessary to power the study adequately.

5) Consider how to handle data truncated due to death.

6) Explore the nature of the missing data in order to inform the missing data analyses.

7) Decide which assumptions about the missing data mechanism are plausible for primary and secondary outcome analyses in light of recommendation 6 (above).

8) Choose and conduct primary analyses that provide valid inferences under the missing data assumptions chosen in recommendation 7 (above), taking into account any auxiliary variables in the model(s).

9) Conduct missing data sensitivity analyses that assess the sensitivity of the results to plausible departures from the primary missing data assumption. These should include an exploration of missing not at random (MNAR) assumptions if plausible.

C. How to report missing data in palliative care trials

To enable users of research to evaluate the risks that missing data pose, clear and complete reporting of data is required. Poor reporting of trials is a persistent source of research waste and considered to be unethical. The box below summarises the recommendations for reporting missing data in palliative and end of life care trials.

A summary of recommendations for reporting missing data in palliative and end of life care trials

In the Methods section:

1) Report strategies used to reduce missing data throughout the trial process.

2) Report if and/or how the original sample size calculation accounted for expected missing data and the justification for these decisions. Report if and/or how the sample size was reassessed during the course of the trial.

3) Report the assumption about the missing data mechanism for the primary analysis and the justification for this choice for all outcomes with missing data.

4) Report the method used to handle missing data for the primary analysis and the justification for the methods chosen, for all outcomes with missing data. Include whether/which auxiliary variables were collected and used.

5) Report the assumptions about the missing data mechanism and methods used to conduct the missing data sensitivity analyses for all outcomes with missing data, and the justification for the assumptions and methods chosen.

6) Report how data that were truncated due to death were handled with a justification for the method(s) (if relevant).

In the Results section:

7) Report the numbers and proportions of missing data in each trial arm.

8) Report the reasons for missing data in each trial arm.

9) Report a comparison of the characteristics of those with observed and missing data.

10) Report the primary analysis based on the primary assumption about the missing data mechanism for all outcomes with missing data.

11) Report results of the missing data sensitivity analyses for all outcomes with missing data. As a minimum, a summary of the missing data sensitivity analyses for the primary outcome(s) should be reported in the main paper with the full results in the supplementary material.

In the Discussion section:

12) Discuss the impact of missing data on the interpretation of findings, considering both internal and external validity.

Introduction

An estimated 85% of research investment is wasted[6]; this equates to approximately $170 billion (approx £126 million) per year[6]. Missing data, in particular, are a persistent source of research waste[7, 8]. They can reduce the power and precision of study findings, as well as how widely the results can be generalised. Crucially, missing data may also introduce **bias**.

To optimise the value of research studies so they may inform decision-making and improve patient outcomes, while optimising the cost-effectiveness of research activity[9], it is essential that healthcare researchers address missing data. For those less familiar with missing data and clinical trials, a plain English overview of how missing data can affect trial results can be found in Appendix 2.

What are missing data?

In this document, the term **missing data** refers to observations (i.e. information) that would be meaningful to help answer a study question and which were intended to be made, but for whatever reason were not observed or documented[10]. This includes data that are missing because the individual providing the data is no longer doing so. This is known as **unit-level missing data**.

Data that are partially missing are also included, for instance when the participant answers some but not all of the questions in a questionnaire. This is known as **item-level missing data**[11]. Furthermore, it encompasses different patterns of missing data, for example when someone completely withdraws from a study and no longer provides any data (known as **monotone missing data**), as well as when participants provide data intermittently (known as **wave missing data**)[11]. Data can be missing due to issues relating to the participant, carers, healthcare professionals, study design, data collector and data input.

Data truncated due to death

Data that are unobserved because a participant has died present a different problem from other types of missing data and do not strictly come under the definition above. We defined missing data above as "**observations that would be meaningful to help answer the study question**", so data that are unobserved because someone has died are not necessarily "missing" at all[12]. For example, if someone has missing data for a quality of life measure and they are alive, the individual still has a certain level of quality of life; we just do not know what it is. In such a scenario, it is appropriate to try and estimate what this may be.

However, if someone is dead, the questions that arise are: how do we define quality of *life* if someone has died – is this ever valid or meaningful? And therefore, should we impute for missing data after death?

To help clarify the distinction between missing data in those alive and those who have died, missing data due to death are often referred to as **data truncated due to death**. We make this distinction in this guidance, although we also recognise that, in terms of quantifying the overall amount and impact of missing data, it can be useful to combine data truncated due to death with missing data in those alive.

Rawpixel/Shutterstock.com

Missing data and palliative and end of life care

Missing data have been shown to be particularly problematic in palliative care trials. A systematic review of 108 palliative care trials estimated that on average 23% (95% confidence intervals 19%, 27%) of primary outcome data were missing at the primary follow-up point[13]. This proportion is much larger than the average found in trials published in major medical journals where approximately 6-10% of primary outcome data were estimated to be missing[14-16].

This may not be surprising given that palliative care patients are, by definition, unwell, often with significant comorbidities and a declining functional ability as their disease progresses. Furthermore, the palliative care trials included in this review were published in a greater variety of journals, and therefore the estimate may be more representative of the amount of missing data across published clinical trials[13].

Such large proportions of missing data can affect the statistical power of a study to detect a difference between treatments.

In palliative care trials that provided sufficient information, 62% (45 of 73) did not achieve the minimum sample size for adequate power specified in their methods section[17]. Moreover, there was evidence that the amount and reasons for missing data differed between the trial arms, suggesting that the missing data may have biased the study findings[13].

Existing guidance about missing data in clinical trials

Several organisations have provided guidance on how to minimise and manage missing data in clinical trials, including The National Research Council (US)[18], the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)[19], and the European Medicines Agency (EMA)[20]. These guidelines have often focused predominantly on the statistical methods to handle missing data and addressed statisticians and methodologists. In 2010, the Methods for Researching End of Life Care (MORECare) collaboration identified missing data as a particular problem in palliative care research and provided some broader guidance based on expert opinion, which they recommended needed further development[21].

Scope of this guidance

This guidance aims to provide an overview of how missing data should be addressed in palliative care trials, building on previous reports. This guidance is focused on randomised controlled trials involving adult participants, though many of the principles will be relevant to other research designs and to paediatric research.

Within this guidance, key considerations are presented in a way that is accessible to all members of the multi-disciplinary team involved in the design, conduct, reporting and implementation of research findings. This includes patients and carers who are interested in the research process, patient and public involvement (PPI) partners, the clinical team, policymakers, commissioners and funders, as well as data collectors, researchers and statisticians.

Each member of the research team is accountable for providing and using robust data, and so it is crucial that all such individuals understand how to reduce, handle and/or report missing data, and the importance of doing so. A list of additional resources containing more detailed explanations is provided at the end of this document.

Development of the guidance

The guidance results from an evaluation of the issues that missing data present to palliative care trials conducted by Dr Jamilla Hussain as part of her **NIHR-funded Doctoral Research Fellowship**. This included:

- a systematic review of trials in palliative and end of life care, which assessed the extent of and factors associated with missing data in this field[13, 22]

- an individual participant-level data analysis of ten palliative care trials, which assessed the participant and site-level factors associated with missing data[23]

- in-depth semi-structured interviews with researchers involved in palliative and end of life care trials with a particular focus on how missing data can be reduced.

All data within this guidance that is referred to as unpublished comes from this work.

In addition, in October 2017 Marie Curie hosted a workshop on missing data in palliative care trials. This involved PPI partners, clinicians, researchers, statisticians and methodologists. During the afternoon delegates helped to refine and develop the guidance on how best to reduce, handle and report missing data in palliative care trials.

How to use this guidance

This guidance is primarily intended to provide PPI research partners, researchers, data collectors, statisticians and research funders and reviewers with a framework for how missing data should be addressed throughout the course of a trial. We hope it will also help interested patients and carers to have a broad understanding of how missing data should be tackled and the importance of complete data provision.

Finally, we hope it will enable clinical team members, policymakers, commissioners and other users of palliative care research to understand the risks missing data may pose to study findings and therefore how and whether study recommendations should be implemented.

The guidance is structured as follows:

- **Part A: how to reduce missing data in palliative care trials**
- **Part B: how to handle missing data in palliative care trials**
- **Part C: how to report missing data in palliative care trials.**

Some of the recommendations are specific to palliative and end of life care; however, many also apply to other areas of healthcare research, especially where participants have advanced disease such as cancer or cardiovascular disease, dementia, multi-morbidity, frailty and when outcomes are participant-reported.

We discuss the recommendations with trials that evaluate clinical interventions in mind; however, the same principles will apply to trials in social care, public health and other fields. The recommendations are based on current evidence and expert opinion, and we aim to update them as new evidence and expertise emerge.

Part A: How to reduce missing data in palliative and end of life care trials

Why reducing missing data matters

To optimise the validity and value of palliative care research, it is essential that missing data are reduced as much as possible. This is because all statistical methods to handle missing data once they have occurred are based on assumptions that cannot be verified[18, 20, 24]. Therefore, reducing missing data in the first place is the most important step in limiting the impact missing data have on study findings. Furthermore, by reducing missing data we can improve the size-efficiency and cost-efficiency of the study, thus reducing waste.

Ethical considerations are an important but under-recognised reason for reducing missing data in research. Patients with palliative care needs want to take part in research as it enables them to contribute to science[25] and leave a legacy[26]. However, they often do this at a point in their lives when both time and energy are limited. So, if researchers do not optimise the opportunities for these participants to provide the data required, this could deprive individuals from contributing to research when this is their wish.

Despite the importance of reducing missing data, most of the advances in the missing data field over the past 50 years have focused on the development of methods and software to analyse missing data once they have occurred. There has been less research on how to effectively minimise missing data in the first place[18].

Factors associated with missing data in palliative and end of life care trials

It is assumed that most of the missing data in palliative care trials are irreversible, i.e. it is due to the patient population being so unwell[21]. However, in a review of palliative care trials, only 26% (1,517 of 5,903) of the instances of missing data were reported as due to disease progression, adverse events or data truncated due to death[22]. This finding suggests that, potentially, a large proportion of the missing data may be avoidable and could be related to the design of research studies.

Studies were more likely to have missing data if they had a longer duration, asked more questions or required more tests[13]. A detailed (individual participant-level [IPD] data) analysis of ten trials in palliative care did, however, find that participants with missing data at a previous time point and those with a poorer performance status (i.e. those who were more unwell) were more likely to have missing data[27].

There was also evidence that factors related to the research site recruiting participants, such as the number of randomisations a site had undertaken and the number of site personnel, were also associated with missing data at the end of follow-up[27]. This study suggests that elements of the process and implementation of trial procedures at the site level may also be important[27].

Addressing each of these participant, trial and site-level factors should be considered when trying to reduce missing data in palliative and end of life care trials.

Evidence on how to reduce missing data

Evidence on how to reduce missing data in other areas of healthcare research is also limited. A Cochrane review of 38 trials testing strategies to improve retention in trials assessed six types of strategies: incentives, communication, new questionnaire format, participant case management, behavioural, and methodological interventions[28]. The review found that monetary incentives versus no incentive had the strongest evidence of effectiveness. Assessment of interventions to improve trial management was, however, limited and most of the interventions were evaluated in single trials in a particular context. The review recommended further trials testing interventions to optimise retention.

In 2018, the James Lind Alliance in conjunction with Trial Forge developed the top ten priorities for research to improve retention in trials[29, 30]. Trial retention overlaps with the issue of reducing missing data. The Priority Setting Partnership report also outlined that further research and development of strategies to reduce missing data in trials are required[30].

Although there is a need for further evidence on how best to reduce missing data in palliative care research, below we summarise the key recommendations based on current understanding and expert opinion. These should be updated as new evidence emerges.

Recommendations for reducing missing data in palliative and end of life care trials

The following are recommendations on how to reduce missing data before and during the trial. Investigators should explain how each of these recommendations will be addressed in the trial protocol.

1) Prepare and plan for how to reduce missing data at the trial design and protocol development stage.

Minimising missing data should be addressed at the start of a trial using a preventative rather than reactive approach.

Optimise the trial design and protocol to minimise missing data.

a) Use a flexible and inclusive study design that facilitates data collection as the physical, psychological and/or social circumstances of the participant change. In particular, consider more than one mode (face-to-face, telephone, post, electronic) and location (home, care home, hospice, hospital) of data collection, and co-develop inclusive strategies to support individuals from diverse backgrounds to participate fully.

Palliative care patients often have an advanced disease that may change, fluctuate and progress during the course of a study. To maximise complete data collection, the likelihood of changes to the physical, psychological and/or social circumstances of the participant needs to be assessed, and appropriate flexibility in how data are collected should be incorporated into the trial design.

There is limited evidence about which methods of data collection are most effective to reduce missing data. Potential strategies may include enabling multiple modes of data collection[31], i.e. using face-to-face consultations, discussions over the telephone, postal questionnaires/surveys and digital means.

Participants are also likely to move between different settings such as home, care home, hospice and hospital. Enabling data collection in all such settings will help to minimise avoidable missing data.

Consideration should also be given to effective ways to optimise inclusive data collection from diverse participant groups such as those with disabilities, living in poverty and/or from minoritised ethnic groups.

b) Consult participants, carers, PPI research partners, clinical teams, data collectors, data managers, statisticians and experienced trialists on how to reduce missing data. In particular, consult experienced data collectors (such as research nurses) for guidance on study design, data collection processes, data volume and case report form design.

Members of the multi-disciplinary team involved in conducting palliative care trials will provide different but potentially important insights into how to reduce missing data in any particular trial.

Feedback on the trial design and data collection process from participants and carers from diverse backgrounds involved in the feasibility and pilot study will provide insights into the participant experience and how to optimise data provision.

Participant feedback will be further facilitated by the expertise of diverse PPI partners who will be familiar with the research methods and study design implications.

Clinical and other care/support teams will provide insight into how they can help support data collection and the potential challenges. Data collectors, such as research nurses involved in similar studies will provide insights into the practicalities of data collection and documentation, including the volume of data requested, and how the trial design can be amended to minimise missing data. Data managers will provide expertise on how to design case-report forms and data input procedures that help to reduce missing data. Statisticians will advise on the data and format of data required.

Experienced trialists will be able to take account of the overall aim and design of the study and practicalities of conducting the study (potentially across different sites), and can help ensure that the design and data will enable the research question to be addressed.

c) Consider how to reduce missing data when specifying the following design features:

- **Research question and primary estimand of interest.**
- **Participants: support all participants to provide data, especially those participants with a poorer performance status.**
- **Duration of the study: use the minimum necessary duration to answer the research question of interest.**
- **Outcomes: define which data are essential to collect given the research question, and only collect these data. Consider outcomes that will minimise missing data. Agree in advance the use of proxies to provide data if the participant is unable to do so. Capture in the case report form whether or not proxies are used or if support is given to participants to complete outcome data.**

Research question and primary estimand of interest

Consider missing data when specifying the research question under investigation and choosing the primary estimand of interest – that is, the numerical quantity that the study is designed to estimate. There are usually several different ways a study can address the specific problem it is investigating, and it is important to formulate a research question that is useful to policy/practice and also feasible to evaluate. This will involve consideration of the risk of missing data.

In terms of the estimands of interest: for example, in an intention to treat analysis, we are interested in the effect of an intervention regardless of whether the participant continues to adhere to the treatment they were allocated[32]. Therefore, participants should continue to be followed up and provide data even if they stop the allocated treatment[33].

Such an analysis closely mirrors clinical practice where patients may not necessarily adhere to treatment and enables the researchers to assess the effectiveness of an intervention in a context closer to real life.

An alternative analysis may be interested in the effect of the treatment only on those who complied with the protocol. An example is per-protocol analysis, which some might consider would not necessarily require a follow-up with participants who did not adhere to the allocated intervention[32].

Even then, further follow-up is likely to be highly informative in attempts to take into account the biases inherent in analysing according to intervention(s) actually received, such as Complier Average Causal Effect (CACE) modelling[34].

The research question being addressed by these analyses are different, and their sensitivity to and risk of missing data will differ[18]. Therefore, missing data should be considered when formulating the research question and primary estimands of interest.

Participants: support all participants to provide data, especially those participants with a poorer performance status

An IPD analysis of palliative care trials found participants with a poorer performance status – i.e. those who are most unwell – are more likely to have missing data in palliative care trials[27]. Yet it is this particular group of participants, with a changing physiology, who may in fact be most vulnerable to iatrogenic harm, i.e. harm caused by the intervention[35].

Therefore, if such patients are, or are going to be, eligible to receive the intervention in practice, it is essential that they are included in the study sample despite the risk of missing data, in order to optimise the external validity of the findings. These individuals will need additional support to provide data, and this should be recognised and addressed as part of the trial design.

Additional barriers to research involvement and data provision may exist for individuals who are marginalised, for example because of their ethnicity, faith, socioeconomic status, refugee status, disability, gender reassignment, sexual orientation, age and/or gender. Equitable support to contribute to trials for these individuals is essential to delivering high quality trials with generalisable findings.

Duration of the study: use the minimum necessary duration to answer the research question of interest.

Palliative and end of life care trials of a longer duration and those that request more data from participants are more likely to have missing data[13]. Minimising the duration and also the individual number of questions asked of, or tests required from,

individual participants should be a principal consideration at the trial design stage.

If the research question of interest and the successful implementation of the findings necessitates an extended length of follow-up and/or more items of data to be collected, and this is justified as part of the protocol, then the risk of missing data needs to be recognised and additional support provided for the participants, carers and research staff to optimise data collection.

Outcomes: define which data are essential to collect given the research question, and only collect these data.

For outcomes where proxy data are valid, at the start of the study researchers should ask the participants if they consent for proxies to provide data if they are unable to during the course of the trial. Proxies can include carers, family members, friends and/or professional staff depending on the outcome. Having more than one potential proxy may help to reduce missing data.

If there are several ways the data may be provided – e.g. independently by the participant, with support from carers, professional or research staff, or from a proxy – this must be captured on the case report form. It is also necessary to clarify whether the data the proxy provides are based on the answer they think the participant would have provided (proxy-participant perspective) or about their own perspective (proxy-proxy perspective)[36].

Even if the use of proxies has not been validated for a particular measure, they can still be used to inform the analyses of incomplete data, i.e. the mechanism of missing data and imputations through their use as auxiliary variables[37] (see introduction to Part B).

d) Evaluate the strategies to reduce missing data during the feasibility and pilot studies. Consider asking participants for feedback on the design and conduct of the study during and/or when they complete or withdraw from the study. Strategies to reduce missing data should also be assessed as part of larger studies potentially within several trials (i.e. a Study Within A Trial) to evaluate their effectiveness, where appropriate.

Strategies to reduce missing data need to be evaluated. For an individual trial, this should form part of the feasibility study where methods to minimise missing data can be assessed with feedback from participants and trial staff on how they found the data collection process and how the trial design may be amended to optimise data collection.

The strategies to reduce missing data should also be implemented as part of the pilot study to assess how well they fit in with the other components of the trial. As part of the main trial, participants and data collectors should be asked for feedback on their experience of taking part in the study and provision of data. This should be during the trial and when participants complete or withdraw from the study. This will help to evaluate and optimise the design of the current study and inform the design

of future studies. These enquiries should be planned from the outset, including securing the necessary ethics and research governance approvals, and reported.

More broadly, to assess the effectiveness of different methods to reduce missing data, it is important that they are evaluated on a larger scale. The evaluation could be part of a Study Within A Trial (SWAT), which is defined as a "self-contained research study that has been embedded within a host trial with the aim of evaluating or exploring alternative ways of delivering or organising a particular trial process"[38]. SWATs also lend themselves to the evaluation of an intervention within several trials to assess its effectiveness and cost-effectiveness across studies[38]. The benefits and role of SWATs are now increasingly recognised, including with major research funders such as the NIHR in the UK[39] and the Health Research Board in Ireland[40].

2) Resource the trial adequately to minimise missing data.

It is imperative that the necessary resources required to reduce missing data are available throughout the course of a trial. Additional funding may be required for such resources, which should be justified as part of the grant application. It is important to recognise that given the bias missing data may introduce, simply inflating the sample size to account for missing data is not sufficient to ameliorate the risk missing data present. Missing data must be minimised in the first place.

Although effective methods to reduce missing data may require further funding, by reducing the risk of bias and improving the efficiency of the trial, these may prove to be cost-effective[18]. There is limited evidence around cost-effective interventions to minimise missing data, and further research is required.

The following suggestions are based on in-depth semi-structured interviews with trial staff involved in palliative and end of life care trials (conducted by Dr Jamilla Hussain from the Wolfson Palliative Care Research Centre)[17].

a) Resource and fund trials adequately to support patients, carers, clinical team members and data collectors to provide complete data. This should include adequate funding for research staff to collect data across different settings and/or using different modalities of data collection, if appropriate.

b) Incentivise sites to provide complete data for each participant or, in the event of incomplete data, the reasons for all missing data. Strategies for this may include regular collective feedback to sites on their performance in this regard, with awards or social media recognition for the best performers. Part of the payment to sites could also be subject to the provision of complete data and/or reasons for all missing data.

Trials often use incentives to encourage sites to recruit participants because poor recruitment is a recognised barrier to trial completion[41]. Similar strategies could be used to promote complete data collection, which will help to raise awareness among site and trial staff that reducing missing data is a key priority and performance indicator.

It is also important to recognise that some level of missing data will be expected and potentially unavoidable. For instance, in a palliative care sample involving participants at the end of life, missing data due to disease progression and data truncated due to death are expected. Such missing data should not necessarily be seen as a sign of poor site performance but rather, perhaps, as an indicator that the sites are recruiting the appropriate participant sample[21]. However, it is essential that when such unavoidable missing data (as well as avoidable missing data) occur, the reasons for the missing data are collected and documented.

Therefore, the key performance indicator with regard to missing data should be for trial staff to try to collect complete outcome data and, if missing data occur, collect the reasons for missing data. See section 5 for further details about collecting the reasons for missing data.

Potential methods to incentivise staff include regular collective feedback to all sites on the proportion of complete data provided and the proportion of reasons for missing data provided per site. This could also involve social media recognition for the best performers at the main data collection time points. Furthermore, it is common practice that part of the funding received by a site is linked to the number of participants recruited. A similar approach where part of the payment is withheld until complete data or the reasons for all missing data are provided may also be effective to reduce missing data.

However, it will be important to avoid unintended negative (i.e. disincentivising) consequences of such an approach. For example, recruiters may exclude participants at risk of deteriorating during the course of the study or with additional barriers to research involvement, even if they meet the inclusion criteria, due to concerns about complete data collection. It is therefore crucial to stress during site training:

- the value of including all participants who are eligible
- the importance of minimising avoidable missing data as much as possible
- if participants are unable to provide data due to reasons that cannot be addressed, then the main aim at this stage is to understand and document the reason(s).

If the reasons are provided in such instances, there should not be a financial penalty. To encourage the collection of all data, however, the incentives to provide complete data should be greater than those to provide incomplete data with the reasons for missing data.

These are only some of the ways sites may be incentivised to reduce missing data, and other examples may be appropriate for different studies depending on the risk of avoidable and unavoidable missing data. Different approaches to incentivise and support sites to reduce missing data should be considered on a trial-by-trial basis and approaches should be assessed as part of the feasibility and/or pilot study, and larger SWATs.

c) Recruit researchers with a good track record for providing complete data and/or reasons for missing data when only partial data are available, and enable them to provide support and mentorship to other researchers.

Mentorship of new sites, or sites less experienced in conducting palliative and end of life care trials, was a central recommendation offered by research nurses involved in palliative care trials, during in-depth interviews[17]. Recruitment of experienced sites and site personnel to take part in a trial and/or to mentor and support less experienced sites could help reduce avoidable missing data at the site-level. This could be facilitated, for example, by face-to-face research practitioner meetings to share best practice across sites, or individual arrangements between sites. Site selection should not, however, necessarily be restricted to those with experience in conducting palliative and end of life care trials to a high standard. By including a mixture of experienced and less experienced sites, trialists can help to build research capacity for future trials and optimise generalisability.

3) Train all research staff to understand the risks posed by missing data and how to minimise missing data.

Training should include:

a) why complete data are important

b) how to communicate with and support participants with palliative care needs and their carers to maximise data collection

c) how to enter data, check for completion and accuracy, and access support for data entry

d) how to document the reasons for missing data.

It is essential that all members of the trial team understand the importance of reducing missing data and their role in addressing this issue. Ensuring research staff prioritise complete data collection at all times requires a good understanding of the risks missing data pose to study validity and generalisability. This requires a clear explanation of the risks of missing data in a way that is understandable to all members of the research team, with regular updates to refresh fundamental concepts.

One of the barriers to data collection for researchers who do not have a palliative care clinical background is feeling apprehensive about asking unwell participants to provide data. Interviews with research staff indicated that those with a palliative care clinical background were more comfortable in supporting participants to provide data, even as they became less well[17].

Training all data collectors on how to approach such situations may help to optimise data collection. PPI partners from diverse backgrounds could be used to develop this training to ensure the approach remains participant-centred and inclusive. To reduce missing data at the documentation stage, training should include the practicalities of entering data, checking it has been entered correctly and how to access timely support when required.

4) Discuss the value of complete data and how to reduce missing data with participants before they consent to enter the trial.

Study participants are also essential members of the trial team who need to understand the importance of both minimising missing data as much as possible and knowing the reasons for missing data.

Discussions should include...

a)...thoughts, ideas and concerns the participants and/or their carershave about the data collection process

Exploration of a participant's concerns and thoughts about the data collection process, before and during the trial, will help to minimise missing data for that particular participant and potentially others. Missing data may be reduced by ensuring the protocol is flexible enough to accommodate any concerns the participants and/or their carershave, or through discussing the rationale behind the study design.

b) ...why each outcome measure is included, the importance of complete data and key data collection time points. This information could be included in the participant information pack in accordance with PPI advice.

A verbal and/or written explanation of why each of the outcomes is included in the trial and why complete data collection is important should be provided to each participant. Evidence suggests that consent information is unbalanced at present, with a greater focus on the rights of participants to withdraw from a study completely without accompanying information that promotes complete data provision[42]. The explanation should be developed with PPI partners and avoid coercive language. A statement could be provided as part of the participant information leaflet/pack or be done separately[42]. Clear, understandable language is essential, and staff should be trained in how to provide this information to individuals from a range of backgrounds.

c) ...why understanding the reason for missing data is important, and asking participants for written or verbal consent to be asked the reason for missing data if this occurs.

Explain to participants that at times missing data will occur and may be unavoidable – this is expected in most trials. However, to understand how to help other participants and trial staff to provide data in the remainder of the trial, and use the data the participant has already provided in a way that will reduce bias, it is essential that the team understands why the data were missing.

Participants should be informed that they can choose not to give a reason, and this will not affect them or the support they receive in any way, but it is important for the team to know that they do not wish to provide a reason and for this to be documented. This is in keeping with the ICH Guidelines for Good Clinical Practice section 4.3.4 which states: "Although a subject is not obliged to give his/her reason(s) for withdrawing prematurely from a trial, the investigator should make a reasonable effort to ascertain the reason(s), while fully respecting the subject's rights."[43]

d) ...a clear explanation that participants can withdraw from parts of the protocol, such as the intervention, but continue to provide data, and why this is useful.

In an intention-to-treat analysis, it is necessary and important that all participants are followed up and their data collected, if this is their wish, according to the intervention they were allocated, regardless of whether the participant continued the intervention as required[32]. Even in the (arguably unlikely) event that the trial only plans to conduct per-protocol analyses, this guidance remains pertinent since such information will be crucial in determining the extent of any bias in such analyses generated through selection effects – whether such investigations involve just descriptive approaches or formal attempts to correct for bias, such as the use of instrumental variables regression methods in CACE analyses[44].

A review of patient information leaflets from trials funded by the NIHR Health Technology Assessment found 70% of leaflets described withdrawal in generic terms and did not make it clear that participants can withdraw from parts of the protocol but continue to provide data[42]. Clarity about this will empower participants to continue to contribute to research if this is their wish.

e) ...asking the participant for written or verbal consent for proxies to provide data about the participant's health or outcomes if the participant is unable to provide this information, and/or for the research team to access their medical records.

Consent at the start of the trial to use proxies and/or the participant's medical records is a useful way to outline how the participant can continue to be part of the study even if the participant is unable to provide direct information themselves.

Discussion about these options may help participants to feel supported to continue in the study even as their disease fluctuates or progresses. Asking for consent also makes it clear that the participant has a choice as to whether proxy information is used and that they can change their mind during the course of the study.

Having the outcome of the consent procedure available to data collectors will enable them to contact proxies or access medical records as soon as possible and therefore limit recall bias.

5) Collect the reasons for missing data.

Understanding the reasons for missing data is central to reducing the extent of missing data, choosing and justifying the approach to analysis and exploring the risk of bias posed by missing data (see Part B section 3b, and Part C section 8). It is, therefore, a core component of tackling missing data.

Despite this, there is evidence that the reasons for missing data are not documented or reported well in palliative and end of life care trials. Over 50% of the reasons for missing data provided in a systematic review of 108 palliative care trials were unclassified and/or uninformative[22]. That is, researchers often used terms such as "lost to follow-up" or "withdrawal" without clarifying what they meant by these

terms or crucially the underlying reason for the loss to follow-up or withdrawal[22].

Knowledge of the reasons for missing data will help trial staff to understand how to reduce future missing data in their current trial as well as other similar studies. The following recommendations outline the key considerations to optimise the collection and utility of this information.

Document the reasons for missing data systematically and clearly:

a) Design a form to collect the underlying reasons for missing data, in collaboration with PPI research partners, clinical team members, data collectors, data managers, statisticians and experienced trialists. Assess and amend the form as part of the feasibility and/or pilot studies with feedback from participants, proxies and the trial team, and continue to review and amend the form as necessary during the trial.

Understanding the reasons for missing data is useful at different stages of the trial, and for different purposes – therefore it is essential that different members of the study team are involved in developing the "reasons for missing data" form. The form should also be assessed as part of the feasibility and/or pilot study and amended as necessary to ensure the information is

clear, informative and usable. During the main trial, the reasons for missing data and issues with data entry should continue to be reviewed and the form amended as required.

b) Include the following reasons for missing data as a minimum: death, disease progression unrelated to the intervention, adverse events or reactions and reasons related to the primary outcome. If the reason is unknown, specify why this was the case.

In palliative and end of life care research, it is important to report the proportion of missing data due to death and disease progression unrelated to the intervention. Both of these reasons are anticipated in most trials that involve participants with advanced disease where the intervention is not expected to affect survival. However, through the randomisation process, on average they should be balanced between trial arms[13]. Collection of these data will, therefore, enable an assessment of the post-randomisation risk of bias (see Part C section 8).

How the judgement of whether disease progression is related to the intervention is made should be explicit in the protocol. This will closely mirror the adverse event reporting process where the principal investigator, having gathered the necessary information, makes an informed judgement as to whether the clinical deterioration is related to the intervention under investigation or not. If it is considered that the deterioration is not related to the intervention and this resulted in missing data, this must be specifically reported as such so it is clear to the reader that this judgement was made.

It should also be explicit who made this judgement and the extent to which they were blind to allocation and treatment. In many cases, especially where this judgement influences the primary outcome variable itself, these assessments should be made by an independent committee/ individual with as much blinding (where certain information is concealed from study participants and/or researchers to protect against bias) as is practicable.

Knowing whether the reason for missing data is related to the primary outcome under investigation will also help to determine the likelihood that the data are missing not at random (MNAR) and therefore which methods to handle the missing data are appropriate (see Part B introduction).

In terms of reducing missing data: missing data due to death, due to disease progression, due to adverse events/ reactions or due to the occurrence of the primary outcome are likely to be unavoidable, but knowing the proportion of missing data due to such reasons will help trialists determine the amount of missing data that is potentially avoidable and where to target their resources.

If the reason for missing data is unknown, the reason for this should be documented. Examples of potential options when documenting unknown reasons for missing data include: participant or proxy not contactable, participant or proxy contacted but unable to provide a reason, data collector unable to provide a reason. Each trial will require different potential options for the reasons for missing data; hence it is important that these forms are developed and piloted before the trial commences and are amended during the trial if necessary.

c) Make a record of who provided the reason for missing data, e.g. participant, proxy, data-collector, clinician, principal investigator, unknown, etc.

Knowledge of who provided the reasons for missing data will help to assess the validity of the reason provided, as well as indicate where confirmatory evidence might be sought if possible.

d) Do not use ambiguous terms such as "withdrawal" or "lost to follow up" without specifying the underlying reason. Avoid the inherently ambiguous term "drop-out" in all circumstances.

Understanding the underlying reason a participant is no longer part of a trial is vital to addressing missing data. Terms such as "lost to follow-up" and "withdrawal" do not provide this information and should be avoided unless they are defined and the underlying reason is provided.

The term "drop out" is often used to cover any or all such eventualities and its inherent ambiguity means that it is best avoided altogether.

6) Distinguish participants who want to withdraw from providing any further data from participants who wish to withdraw from part of the study protocol but consent to ongoing data collection or access.

a) Clarify whether a participant who wants to withdraw from the trial seeks to (i) withdraw from the intervention fully or partially, or (if possible) continue the intervention; (ii) withdraw from data collection fully, partially or not at all; or (iii) withdraw from the use of proxies or their medical records to gather data. Continue to follow-up participants who do not withdraw from all data collection specifically, even if they discontinue the intervention fully or partially.

b) Ask those who withdraw from all data collection the reason for this and, if possible, ascertain the primary outcome at the time of complete withdrawal.

Whenever a participant requests to withdraw from the trial, it is important to clarify whether they want to withdraw from the intervention, withdraw from data collection and/or withdraw from the use of proxies or their medical records (see 4d).

The extent to which they want to withdraw from these components of the protocol also should be established, i.e. fully, partially or not at all.

If the participant opts to withdraw entirely from the intervention, data provision and use of proxies or medical records, it is important to ask the participant the reason for this and, if possible, the primary outcome at this time, which will inform the analysis of the data they have provided to date. The participant has the right not to provide this information if this is their wish and this should be documented.

7) Monitor and address missing data during the trial.

a) Have alerts when missing data occur, especially for primary outcomes. Clarify whether this is missing data or delayed data collection to ensure alerts are accurate.

b) Monitor the following as key performance indicators throughout data collection:

- **The number and characteristics of participants with missing data in each trial arm**
- **The reasons for missing data and the number of participants with missing data for each reason, in each trial arm**
- **The number of participants with missing data without a documented reason, especially for the primary outcome, in each trial arm.**

During the trial, it is essential to monitor the amount of and reasons for missing data in order to identify issues with incomplete data early. Knowing the reasons for missing data will help trialists understand whether or how future missing data can be prevented and will allow problems relating to data collection or documentation to be addressed acutely.

If the protocol is flexible, it is important to distinguish between data that are missing from data being collected later and/or from another source.

Monitoring the rates of missing data in each trial arm will help highlight whether there is a difference between arms that may introduce bias.

Tracking whether the reasons for missing data are documented will enable any issues with the collection or documentation of this information to be identified early and addressed.

Part B: How to handle missing data in palliative and end of life care trials

Even with the best efforts to reduce missing data, some level of missing data is expected in most trials. It is important that datasets with missing data are handled in line with established principles. This necessitates that the methods chosen for analysis are based primarily on plausible assumptions about the missing data mechanism and not computational ease.

Principles of missing data analysis

In this section, an overview of the key concepts and components of a missing data analysis is provided. It aims to help those unfamiliar with missing data analyses to understand the rationale for the recommendations for handling missing data.

Statistical analyses of trial data often involve fitting a statistical model to assess the relationship between the outcome variable of interest (i.e. the outcome measured during the study to assess the impact of a given intervention) and potential explanatory variables (i.e. factors that may explain the change in the outcome variable). In this guidance, we refer to the statistical model assessing this relationship as the main model and distinguish between the outcome and explanatory variables.

We will use the example of a trial testing a treatment for pain, where a pain score is the outcome of interest.

What are the mechanisms of missing data?

Little and Rubin (2002) describe a commonly used taxonomy for missing data based on the mechanism for missingness[45]. They suggest that data can be:

- **Missing completely at random (MCAR):** where the missingness is nothing to do with the person being studied and the chance of having missing data is the same for all participants and represents a completely random process[3].

For example, data are MCAR if a pain outcome is measured through a postal questionnaire and the questionnaire was lost in the post due to an unsystematic error. For example, the postal worker dropped it and therefore did not deliver it. This incident has nothing to do with the person being studied and such random events should, on average, balance out between trial arms.

As the distribution of missing values is expected to be the same as the distribution of observed values, data that are MCAR are not expected to result in bias under common analysis methods[45].

- **Missing at random (MAR):** where the missingness (a) is related to the person being studied but it can be predicted from other information about the person[3]and (b) is not specifically related to the information that is missing[3].

For example, data are MAR if (a) the pain score is more likely to be missing in participants who are older and their age is known and (b) the pain score is not missing directly due to pain.

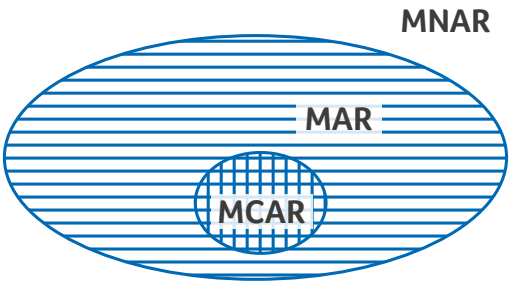
MAR data are related to observed variables, but not to unobserved variables (such as the pain score), and therefore are not expected to lead to bias if handled appropriately. Here the recorded variables about the person can account for any differences between the distributions of observed and missing values[45]. If data are MCAR, then they are also MAR.

- **Missing not at random (MNAR)**: If the data are not MCAR or MAR, they are MNAR. This is the case when missingness is specifically related to the information that is missing[3].

For example, if the missing primary outcome is a pain score, data are MNAR if a participant's data are missing due to their pain.

The recorded data in this instance do not account for the differences between the distributions of observed and missing values[45]. Thus, the missing data depend on unobserved data. Data that are MNAR result in bias under common analysis methods and are difficult to analyse correctly.

Figure 1. Taxonomy for missing data



MNAR: Missing not at random

MCAR: Missing completely at random

MAR: Missing at random

Deciding on the mechanism

It is not possible to know for sure, based on the data that are observed, what the underlying mechanisms are for the missing data[46]. This is because we cannot be sure if the mechanism is related to the missing values – that is, we cannot be sure that the data are not MNAR.

So, it is imperative to carefully consider the possible plausible mechanisms based on both the data and, crucially, the documented reasons for missing data, while bearing in mind that these cannot be verified using the observed data.

Often this means doing a main analysis in which a simpler assumption is made (often MAR) and a sensitivity analysis in which more complex assumptions are made (often MNAR). If MAR is not plausible, however, the primary analysis should include analyses that allow for MNAR.

Which method(s) should I use to handle missing data?

There is no single method to analyse missing data that is recommended for all trials. The method(s) chosen should be based on the assumed mechanism(s) of the missing data[47]. It is also highly desirable to use all the data provided. This is often useful from a statistical perspective, but also important ethically to enable as many participants as possible to contribute to research findings.

a) Missing completely at random (MCAR)

If the data are assumed to be MCAR, it is often argued that **complete case analysis** is potentially a valid method of handling the missing data[46]. This method excludes participants with missing data in at least one variable in the main model, from the analysis. It is usually the default method used to handle missing data in statistical software.

An MCAR assumption suggests that the complete data are a representative sub-sample of the original sample, and that therefore limiting the analyses to this subset of participants will still provide valid estimates[48].

However, it is likely to be rare that such an assumption is plausible in palliative care trials. Moreover, complete case analysis is inefficient, wasteful and potentially unethical as it ignores the data that was provided by individuals who wanted to contribute to a trial but were unable to provide complete data[10]. For these reasons we recommend using the methods discussed under MAR below, even if data are believed to be MCAR.

b) Missing at random (MAR)

For data that are assumed to be MAR, simple imputation methods, where a single value is imputed for the missing data and then the dataset is treated as if it were the fully observed data, are generally discouraged[10]. Such methods do not account for the fact that we cannot be sure about the true values of the missing data; these have been shown to produce biased estimates[24]. Examples of these include:

- **last observation carried forward –** where it is assumed that the missing response is the same as the last observed response
- **baseline observation carried forward** – where it is assumed that the missing response is the same as the baseline response
- **mean imputation** – where the missing value is replaced by the average of the observed values for that variable
- **regression mean imputation –** where a single predicted mean is imputed based on the regression of the incomplete variable on the complete variables[10, 48].

Mean imputation and regression mean imputation are, however, valid – indeed, recommended[49] – ways to handle an incomplete explanatory variable in a randomised trial when estimating the treatment effect. However, they are not valid in observational studies and should never be used for outcomes[49].

Alternatively, more principled methods of analysis under an MAR assumption, such as **multiple imputation, model-based methods and inverse probability weighting**, do not attempt to replace the missing values directly, but instead use observed data to generate statistical information about the missing data and/or their mechanism[10, 18-20].

In **multiple imputation**, as the name suggests, multiple imputations are created for the missing data. It involves three stages:

1) Imputing the missing data a number of times (typically denoted as m times): the imputations are based on the observed data using algorithms, and several datasets with different imputations are created[50]. The imputed values are sampled from a predicted distribution based on the data provided[5]. The variability between the datasets, introduced by the imputation procedure, reflects the uncertainty in estimating the missing values and this variability is used to increase the estimates of precision[10, 50].

2) The m imputed datasets are analysed separately using standard statistical procedures[50].

3) The estimates from each imputed dataset are combined using a method that takes account of the variability, i.e. the difference between imputations as well as within each dataset (this is known as Rubin's Rules)[10].

Model-based methods do not replace the missing values explicitly, but use an algorithm to generate parameter estimates (e.g. treatment effects at different time points) that take account of the observed data, missing data, the relationships among the observed data and assumptions about the underlying distribution[51].

One common model-based procedure is known as maximum likelihood, which chooses parameter estimates for the model with values that maximise the likelihood of obtaining the observed data[10, 52]. When data are missing, and the mechanism is MAR, the likelihood can be obtained by estimating probabilities across all possible values of the missing data[46].

The method of maximum likelihood handles missing outcome data easily and is therefore usually reserved for this case[48]; it is much more complicated to handle missing explanatory variables.

Model-based methods can be extended for data that are assumed to be MNAR by including an explicit model for the missing data mechanism[48].

Inverse probability weighting involves conducting a complete case analysis; however, different weighting is given to different participants with complete data, depending on how similar they were to those with missing data[48]. Participants with complete data who were most similar to those with missing data, i.e. they had a high probability of having missing data, are given more weighting in the analysis, than those who had a low probability of missing data[48.] This is to make the sub-sample of complete cases more representative of the population of interest.

Finally, when the outcome is measured at one time point (rather than repeatedly over time), **complete case analysis** is valid under an MAR assumption, provided that the regression model includes any variables that predict both missingness and outcome[48].

Indeed, in this instance, complete case analysis is preferable to more complex analyses[53]. However, if there are concerns about the representativeness of participants with complete data, out of all participants randomised, then multiple imputation within trial arms or inverse probability weighing is recommended.

c) Missing not at random (MNAR)

For MNAR data, it is important to determine the underlying reason for missing data. For example, if the primary outcome is a measure of pain, it would be useful to find out if the data were missing because the participant at this time had too much pain, the pain had resolved or the pain had not changed. Often this is not known, and it is not possible to determine whether this is the case from the observed data.

It is possible to use **multiple imputation** and **model-based methods** to analyse MNAR data[46]. However, a model for the missing data mechanism needs to be specified[10, 54]. There will be numerous ways the missingness may be associated with the unobserved data, so it is essential to assess the sensitivity of the estimates to different MNAR models in a sensitivity analysis[33]. In particular, the flexibility to adjust imputed values in multiple imputation lends itself well to conducting such sensitivity analyses[24].

Discussion between statisticians and other members of the trial team are necessary to determine a plausible range of values for this adjustment: for example, imputed values might be increased by 5%, 10% and 15% in three separate analyses, where a difference of 15% between missing and

observed values is the largest difference considered plausible.

What about data truncated due to death?

Analysis of data truncated due to death may require different statistical approaches. How to handle such data is a developing area of research, and several methods have been described[12, 55-57]. The approaches can be categorised according to how or whether the analysis takes account of survival[55].

Potential options for analysis may include ignoring survival and imputing the data as if no one died[55]. In many cases this would not be appropriate in palliative care research where outcomes such as quality of life are often used, and quality of life after death is not a meaningful measure.

Another approach is only to consider participants who survive, but this would give an effect estimate just for those well enough to survive to the primary follow-up point and would not be generalisable to those who did not[58]. In palliative care research, it is often important to understand whether the treatment benefits or harms those who are deteriorating, as these individuals often receive the intervention in clinical practice until the point of death. Therefore, it is important to consider methods that include participants who do deteriorate and die[58].

Possible strategies are to use a composite endpoint where mortality and outcomes at all points before death are combined, for example using a utility-based measure where the utility assigned to death is defined; or to perform survivors-only analysis at each time point and then estimate the average treatment effects over time[58].

Each of these approaches to analysing data truncated due to death addresses different research questions. It is therefore essential that the method(s) chosen are carefully considered to ensure the estimate is relevant to practice and applicable[1]. The findings of different approaches may also be compared in supplementary analyses[1].

What is the role of missing data sensitivity analyses?

No matter what method is used to handle missing data in the primary analysis, it is essential to conduct a sensitivity analysis that assesses the sensitivity of the findings to different assumptions about the missingness mechanism[18, 20, 33, 59]. This is necessary because all methods to analyse missing data are based on assumptions that cannot be verified from the (partially) observed data[18, 54].

A principled approach to a missing data sensitivity analysis would require that the assumptions about the missing data mechanism are varied systematically[33]. For example, if the assumption of the primary analysis is MAR, then the sensitivity analysis should instead assume that the missing data are not MAR (i.e. MNAR) and would specify, in a systematic fashion, a range of ways in which the missing values may differ from the imputed values[33]. The departures from the primary assumption should be plausible and informed by the reasons for missing data and clinical expertise.

How well do palliative and end of life care trials handle missing data?

In a review of 108 palliative and end of life care trials, 93 reported some missing data[13]:

- Only 3% (3 of 93) reported their assumed mechanism of missing data[22].
- 60% (56 of 93) of trials with missing data conducted complete case analysis alone to handle missing data[22].
- 23% (21 of 93) used multiple imputation or model-based methods.
- Only 16% (15 of 93) reported a missing data sensitivity analysis[22].

In comparison, a systematic review that assessed how missing data were handled in trials published in five major general medical journals in 2001 found 21% of trials with missing data reported a missing data sensitivity analysis, and 45% used complete case analysis as their primary analysis[15]. A review of the same journals in 2014 found that 35% reported a missing data sensitivity analysis and 45% used complete case analysis as their primary analysis[16]. Thus, palliative care trials (albeit published in a greater variety of journals) were less likely to use a principled approach to handling missing data.

There is a possibility that the trialists in these studies conducted further analyses based on the above principles but did not report them due to word count limitations. However, a review of missing data in 517 Health Technology Assessment monographs, which do not have a word-count limit, found missing data were rarely discussed in the missing data section of the report and missing data sensitivity analyses were only reported in 30% of trials[60]. This review suggested missing data are not handled appropriately even in what could be regarded as a sample of well-resourced trials.

Overall, the evidence suggests there is a need to improve the handling of missing data across all trials. To facilitate this, clear guidance on the key components of handling missing data is required.

Recommendations for handling missing data in palliative and end of life care trials

Below are the recommendations for a principled approach to handling missing data in palliative and end of life care trials based on current evidence and expert opinion. A brief explanation is provided for the included criteria, with references for resources that provide more detailed information available in the **Further Reading and Resources section** at the end of this document.

1) Include a statistician in the trial team during the design, conduct and analysis stages of the study.

Statisticians can provide expert advice on how to optimise the trial design, conduct, analysis and interpretation of findings in order to minimise the impact of missing data. They should be part of the trial team from the beginning and provide specific guidance on each of the following recommendations.

2) Decide how missing data will be handled in the design and conduct of the study and in its analysis. Report these decisions in the protocol and statistical analysis plan.

In the protocol, address recommendations 3 and 4 below in particular, which refer to missing data handling considerations at the design stage. In the statistical analysis plan, address in detail recommendations 5-9, which refer to how missing data will be analysed.

3) Prepare for missing data analyses at the trial design stage.

a) Reduce missing data as much as possible (as per the guidance on reducing missing data in Part A).

Missing data reduce the information available and complicate statistical analysis. Reducing missing data will therefore minimise the risk of bias posed by missing data analyses.

b) Collect the reasons for missing data systematically and clearly, so the information collected can be used to inform the choice of plausible assumption(s) about the missing data mechanism and which methods to handle missing data may be appropriate.

Work with statisticians to design forms to collect the reasons for missing data. Include a broad range of reasons that can be categorised and are informative at the analysis stage.

Include the following reasons as a minimum: death, disease progression unrelated to the intervention, adverse events or reactions, and reasons related to the primary outcome.

If the reason is unknown, specify why this was the case.

Assess and, if necessary, amend the forms as part of the feasibility study.

Reasons for missing data help to inform decisions about the assumed mechanism(s) of missing data and therefore which method(s) should be used for the primary and sensitivity analyses. For the documented reasons to be informative and usable in this way, the potential options for the "reasons for missing data" form must be comprehensive and categorised in a way that will aid the judgement about the missing data mechanisms. For example, if the reason states that data were missing due to:

- **a completely random event unrelated to the participant, trial or data**, this makes an MCAR assumption plausible. In such instances, detailed information should be provided about the event, so the judgement as to whether the reason was related to the trial can be made by the trial team as a whole and not just the data collector

- **disease progression, which was monitored, for example, using performance status data**, this makes an MAR assumption plausible

gorodenkoff/iStockphoto.com

- **the primary outcome of interest, which was missing**, an MCAR/MAR assumption is therefore not reasonable. How the missingness is likely to be related to the primary outcome will help to inform the plausible adjustments to the imputations for the missing data sensitivity analysis. It may not often be possible to determine if missingness is potentially related to the primary outcome, but the participant, proxy or data collector may be able to provide information that helps to reduce the uncertainty.

Furthermore, including death as a reason is important because the approach to handling data truncated due to death may differ from that of missing data in participants who are alive. See Part A section 5 for further recommendations on collecting the reasons for missing data.

c) Consider whether any auxiliary variables should be collected.

These may include:

- **the Australian-modified Karnofsky Performance Status (AKPS)**
- **other clinical variables that predict missingness or key incomplete variables**
- **the outcome variable of interest measured at an intermediate time point.**

Auxiliary variables are variables not in the main model, but associated with the missing data[45]. They can be used in imputation models to strengthen the plausibility of an MAR assumption, reduce bias and improve statistical power when missing data occur[45, 61]. Auxiliary variables should be carefully selected based on theory and evidence and should be highly correlated with the variables with missing data[61]. They are collected in addition to the other variables in the analysis and therefore the burden of additional data collection, and thus increased risk of missing data, should be weighed up against the benefits for the analysis.

The AKPS was found to be strongly associated with missingness in an individual participant-level data analysis of ten palliative care trials[27]. This finding indicates it may be a useful auxiliary variable to include in the imputation model (step 1 of multiple imputation – see Part B introduction) in trials where missing data due to disease progression are expected.

An analysis of palliative oncology trials in one centre in the United States also found some evidence that higher intensity fatigue and breathlessness, Hispanic race, higher education level and religious affiliation were associated with missingness[62].

Further research is required to develop the evidence base for auxiliary variable selection in palliative care trials. The examples given are suggestions based on current evidence and expertise. Trialists should consider in advance which auxiliary variables will be most valuable to collect for their particular trial and evaluate them as part of the feasibility and pilot studies.

4) Inflate the sample size to account for expected missing data in order to achieve the number of participants necessary to power the study adequately.

a) Decide on the appropriate sample size for the study (without missing data).

b) Estimate the proportion of missing data expected for the primary outcome and justify this using one or more of:

- **a feasibility study of some kind**
- **similar studies elsewhere**
- **evidence of factors associated with missing data in similar studies**
- **expertise in the field, e.g. clinical knowledge.**

c) Inflate the sample size so that the expected amount of observed data is the appropriate sample size (e.g. if 20% missing data are expected, the sample size needs to be multiplied by 1/[(100-20)/100], which is a 25% increase).

d) Re-evaluate the trial procedures, feasibility and sample size during the trial if the extent and/or nature of missing data is substantially different from that anticipated. Use pre-defined rules to reassess sample size.

Missing data can reduce the power and precision of trial findings. The final sample size required for a trial is usually calculated by first determining the minimum sample size needed to adequately power the study and then inflating this by the proportion of missing data expected. Increasing the sample size in this way will help to address the loss of power, but it will not address the risk of bias posed by missing data.

In a systematic review of over 100 palliative care trials, only 62% (45 of 73) achieved the minimum sample size estimated as required once missing data was taken into account[17]. This finding suggests these trials under-estimated the proportion of missing data. It is therefore important that the predicted proportion of missing data is based on evidence, such as previous experience of similar trials or the feasibility/pilot study. Evidence suggests that trials with a longer duration[13], those that request more data[13] and include participants with a poorer performance status[27] are more likely to have missing data in palliative care trials – these characteristics of the study design should be used to inform the sample size calculation.

5) Consider how to handle data truncated due to death.

It is usually inadvisable to use methods that impute values after death in palliative and end of life care trials, as the values of the outcome if death had not occurred are not meaningful for clinical practice. Consider:

a) survivors-only analysis when the proportion of participants remaining alive is not anticipated to be affected by the intervention. Such an assumption must be justified

b) composite approaches that combine survival and the outcome when there are differential rates of data truncated due to death.

See Part B introduction. Resources with further explanation and discussion about handling data truncated due to death are available in the **Further Reading and Resources** section.

6) Explore the nature of the missing data in order to inform the missing data analyses.

a) Tabulate how much missing data there is for each variable by trial arm.

b) Explore patterns of missing data.

c) Tabulate the reasons for missing data for primary and secondary outcomes by trial arm and discuss any queries with the data collector, clinical team, data managers and/or PPI research partners.

d) Compare the distribution of major variables between participants with the outcome missing and those with it observed, and explore whether observed variables predict data being missing.

The first step of the analysis of datasets with missing data should be to describe and explore the data. This should include tabulating how much missing data there is for both the outcomes and any explanatory variables by trial arm to determine the risk posed by missing data. If there are different rates of missing data in each trial arm, this may indicate that the missing data have introduced bias[63].

Exploring the pattern of missing data may also be informative. Typically this involves determining whether the data are missing in a monotone pattern, i.e. when the participant has missing data at one time point and all subsequent data are also missing[48]. This pattern usually occurs in trials where participants are followed up over time, and the participant completely withdraws from the trial and all data collection. Identifying such a pattern of missing data may help to inform decisions about the reason and mechanism of missing data, and simplifies the use of inverse probability weighting and multiple imputation[48]. Assessing the pattern of missing data for auxiliary and outcomes variables also helps to determine the value of the auxiliary variables – for example, if the auxiliary variable is frequently collected when the outcome is missing. This pattern indicates that the collection of the auxiliary variable is useful.

Understanding the reasons for missing data per trial arm will help to determine whether there are differential reasons for missing data in each arm. This may suggest that different types of participants have missing data in each trial arm and therefore that the missing data may introduce bias unless this is accounted for in the analysis. If there are data truncated due to death, a comparison of the baseline characteristics of those with data truncated due to death with those with missing data for other reasons may also be useful.

The observed data can also be used to inform decisions about the mechanism of missing data. For example, if there is evidence of a substantial difference in the distribution of variables according to whether the outcome variable is observed,

this provides evidence against the assumption that the outcome variable is MCAR[9]. Regression analyses can be used to assess the association between missingness of a particular variable and other variables[9]. However, it is important to note that such analysis cannot confirm or refute the underlying mechanism[9].

7) Decide which assumptions about the missing data mechanism are plausible for primary and secondary outcome analyses in light of recommendation 6 (above).

Use the reasons for missing data and exploration of the data to inform the positing of plausible assumptions about the missingness mechanism and decide upon the assumption(s) for the primary and secondary outcome analyses. Generally, an MAR assumption is considered more realistic than MCAR (see introduction to Part B).

8) Choose and conduct primary analyses that provide valid inferences under the missing data assumptions chosen in recommendation 7 (above), taking into account any auxiliary variables in the model(s).

The plausible assumptions about the missing data should inform the methods used to handle missing data. Additional considerations when choosing between different valid approaches include how much data are missing, which variables are missing, the pattern of missingness and computational efficiency[9] (see introduction to Part B).

9) Conduct missing data sensitivity analyses that assess the sensitivity of the results to plausible departures from the primary missing data assumption. These should include an exploration of missing not at random (MNAR) assumptions if plausible.

The assumptions about the missing data mechanism cannot be verified using the data that are observed. Therefore, it is important to assess the sensitivity of the findings to different assumptions about the missing data mechanism by performing a number of different sensitivity analyses that are valid under different assumptions[33] (see introduction to Part B).

Part C: How to report missing data in palliative and end of life care trials

To enable research users to evaluate the risks missing data pose, transparent and complete reporting of data is required. It has long been recognised that all trials need to be reported clearly.

DerSimonian and colleagues stated that "editors could greatly improve the reporting of clinical trials by providing authors with a list of items that they expected to be strictly reported"[64]. This recognition spurred the creation of the Consolidated Standards Of Reporting Trials (CONSORT) statement, which specifies the items of information that must be reported in order to assess the internal and external validity of trial results[32]. Missing data present a substantial risk to the validity of trial findings and can introduce bias. Therefore, transparent reporting of missing data is paramount[65].

CONSORT 2010 statement

The most recent CONSORT 2010 Explanation and Elaboration report considers the impact missing data may have on the validity of intention to treat analyses, although missing data are only explicitly addressed in the results section[32]. The checklist includes recommendations to report a participant flow diagram (item 13) and to report for each group the numbers of participants that were randomly assigned, received the intended treatment, and were analysed for the primary outcome (item 13a).

Rawpixel/Shutterstock.com

Furthermore, it states that losses and exclusions after randomisation, together with the reasons, should be reported (item 13b). It also recommends that the number of participants included in each analysis be reported, as well as whether the analysis was by the original assigned groups (item 16).

However, specific guidance is not given on whether or how to report the assumptions about the missing data mechanism, the methods of handling missing data, and a missing data sensitivity analysis[22]. Furthermore, there is no specific recommendation that the influence of missing data on the findings should be reported as part of the discussion[22].

It could be argued that the application of the general CONSORT principles would encourage the reporting of all of these[66]. However, a systematic review of the reporting of missing data in palliative care trials found that each of the missing data reporting criteria specified in the CONSORT 2010 statement were reported by at least 69% of trials; those not specified by CONSORT were less well reported[22].

As discussed previously, only 3% (3 of 93) of trials reported the assumed mechanism of missing data, 48% (45 of 93) reported the methods used to handle missing data, 5% (3 of 60) reported methods used to handle data truncated due to death, and 16% (15 of 93) reported any sort of missing data sensitivity analysis[22].

Additional guidelines

Other guidelines have recommended additional criteria for reporting missing data. Akl et al conducted a systematic survey of the literature to identify recommendations for reporting missing data in trials[67]. They selected 13 articles including the CONSORT 2010 statement for parallel trials[32] and the CONSORT extensions for patient-reported outcomes[68], harm[69] and cluster trials[70]. From these recommendations, they provided reporting guidance covering the proportion, reasons, patterns, analytical methods and interpretation of missing data[67].

The MORECare collaboration proposed that palliative care trials report the proportion of attrition due to death and illness specifically[21]. Dumville et al. encouraged trialists to report a comparison of baseline characteristics between trial arms for participants included in the analysis and separately for those with missing values[71].

These comparisons aim to assess the impact of missing data on the balance between trial arms for measured participant characteristics, although it is important to note that trials are not necessarily powered for such tests and the absence of differences does not guarantee lack of bias[22].

The National Research Council recommended that sensitivity analyses should be "a mandatory component of reporting"[18]. A systematic review and Delphi survey that focused on missing data in patient-reported outcome research specified that the potential influence of missing data on the findings should be reported[65].

Reporting of missing data in palliative and end of life care trials

A systematic review of the reporting of missing data in palliative and end of life care trials found 69% (75 of 108) of trials accounted for all participants who entered the study, 94% (101 of 108) reported the number of participants not included in the primary outcome analysis and 87% (85 of 98) reported the number of participants with missing data in each arm in the primary outcome analysis[22].

However, of the 99 trials with secondary outcomes, only 9% (9 of 99) reported the proportion of missing data for all secondary outcomes and 18% (18 of 99) for some secondary outcomes[22]. Of those with repeated measures, only 7% (5 of 69) reported the proportion of missing data at each time point, and those with primary outcomes that were scale summaries, only 10% (5 of 50) reported the amount of item-level missing data[22].

Thus, the proportion of missing data reported in palliative and end of life care trials is likely to be an under-estimation of the true proportion, and better reporting is required. Although 71% (66 of 93) reported the reasons for missing data, over 50% of the reasons were uninformative[22].

Of the trials that reported missing data in the systematic review, 60% (56 of 93) used complete case analysis alone to handle data, yet only 6% (6 of 93) reported a comparison of baseline characteristics between trial arms of those with observed data, and no trials reported a comparison of those with missing data[22]. Only 16% (15 of 93) of palliative care trials with missing data reported the use of more than one method to handle missing data, of which 11 trials reported the complete case analysis as the primary analysis without justification for this choice[22].

Less than half (43 of 93) of palliative care trials with missing data discussed the potential impact of missing data on the interpretation of the trial results and few specifically outlined the impact on the internal and external validity of the study[22].

The need for further guidelines

To improve the reporting of missing data in palliative and end of life care trials, and therefore the usability of research findings, clear and comprehensive guidance on the key criteria that must be reported is required. The recommendations below are based on the results of the systematic review of missing data reporting in palliative and end of life care trials[22] and build on, and bring together, the recommendations by CONSORT[32] and other groups[18, 21, 65, 67, 71].

The focus of the guidance is on the reporting of the original publication in peer-reviewed journals and research reports, but the criteria should be considered in all forms of communication of the findings. They should act as a guide not only for trialists reporting their study, but also for peer reviewers, journal editors, funders and research users such as patients, clinicians, policymakers and commissioners to enable them to assess the risk posed by missing data and the methods of analysis.

Recommendations for reporting missing data in palliative and end of life care trials

The recommendations for reporting missing data are categorised according to the information that should be reported in the methods, results and discussion section of a publication or report. The rationale for the approach to reducing and handling missing data are covered in the recommendations in parts A and B of this report. These recommendations focus on criteria that must be reported to enable research users to make an informed judgement about the risk posed by missing data, and the methods of analysis, to the validity of the study findings.

Methods section

1) Report strategies used to reduce missing data throughout the trial process.

Reporting the strategies employed to reduce missing data will help readers to make a judgement of how well missing data were reduced. It may also help other trialists consider how to optimise complete data collection.

2) Report if and/or how the original sample size calculation accounted for expected missing data and the justification for these decisions. Report if and/or how the sample size was reassessed during the trial.

If the estimated sample size required was inflated to account for missing data, this should be reported. The proportion of missing data expected and justification for this estimate should be specified. If the sample size was reassessed during the trial due to the extent and/or nature of missing data, this should be reported along with the justification for the reassessment. See Part B section 4.

3) Report the assumption about the missing data mechanism for the primary analysis and the justification for this choice, for all outcomes with missing data.

The choice of methods to handle missing data should depend on the assumptions about the missing data mechanism. The assumptions should therefore be specified to enable research users to evaluate whether the methods used were appropriate. Different assumptions may be made for different outcomes, and these should be reported separately. If secondary outcome analyses are only provided in the supplementary material, the missing data assumptions (and methods to handle the missing data) for these outcomes could be provided as supplementary material only. See Part B sections 6 and 7.

4) For all outcomes with missing data, report the method used to handle missing data for the primary analysis and the justification for the method chosen. Include whether or which auxiliary variables were collected and used.

It is necessary to clearly specify the method used for the primary analysis, even if this was the default method of complete case

analysis. Complex methods of analysis typically require detailed description. If auxiliary variables are used for multiple imputation, these should be reported; if no auxiliary variables are used, this should also be stated.

5) Report the assumptions about the missing data mechanism and methods used to conduct the missing data sensitivity analyses for all outcomes with missing data, and the justification for the assumptions and methods chosen.

Both the assumptions about the missing data mechanism and the methods used to assess the sensitivity of the findings to different assumptions about the missing data mechanism must be reported in the methods section. This is so the reader can assess the robustness and relevance of the findings.

6) Report how data that were truncated due to death were handled with a justification for the method(s) (if relevant).

If there are data truncated due to death, the methods used to handle this data must be reported in the main report with the justification for the method(s) chosen. See Part B section 5.

Results section

7) Report the numbers and proportions of missing data in each trial arm.

Report the following for each outcome, at each time point the outcomes are measured, and separately for each trial arm:

a) Number of participants who have died.

b) Number of surviving participants in each arm with missing data (unit-level missing data).

c) For outcomes that are scale summaries (such as questionnaires and surveys): the amount of item-level missing data – for example, the number of participants with at least some items of the questionnaire missing, and the proportion of item-level missing data.

The proportion of missing data must be reported clearly and fully. This includes reporting the proportion of missing data for both primary and secondary outcomes, at each measurement time point for outcomes measured repeatedly over time, and for both unit and item-level missing data.

Reporting the proportion of missing data per trial arm will allow the reader to assess whether there is an imbalance between the trial arms, and therefore whether the missing data may have introduced bias post-randomisation.

Data truncated due to death presents a different problem in terms of how to handle missing data (see Part B introduction), it is therefore useful to separate this from missing data in those alive and report the proportions separately.

8) Report reasons for missing data in each trial arm.

a) As a minimum, report the number of participants with data truncated due to death and those with missing data due to disease progression unrelated to the

intervention, adverse events or reactions, and reasons related to the primary outcome.

b) If terms such as "lost to follow up" or "withdrawal" are used, these must be defined and the underlying reason reported.

c) If different reasons for missing data are provided for different outcomes, these should be reported separately.

Evidence of different reasons for missing data in each trial arm suggests that different types of participants may have had missing data in each trial arm and therefore that the missing data may have introduced bias.

Knowledge of the reasons for missing data will also enable the reader to make a judgement about the external validity of the study findings. For instance, if a large proportion of data are truncated due to death and disease progression, and these participants are not included in the analysis, the results may not be generalisable to these patients who are less well and ill enough to die.

The underlying reasons for missing data (and the variables collected) will also help the reader assess whether the assumptions about the missing data mechanism were plausible. For example, if data are missing due to disease progression and performance status data are collected, then it may seem reasonable to make an MAR assumption (see Part B introduction).

Furthermore, knowing the proportion of data truncated due to death will enable the reader to make a judgement about the risk posed by the method chosen to handle data truncated due to death. For example, if very little data are truncated due to death (such as <5%) with similar proportions in each trial arm, there may be less concern about the risk posed by the method of analysis, compared with trials that have a large proportion and/or differential rates of data truncated due to death across trial arms.

In trials where data truncated due to death or there is missing data due to disease progression, adverse events/reactions or related to the primary outcome are a possibility, the proportion of missing data due to these reasons must be provided as a minimum. This is the case even if none of the participants were missing due to these reasons.

It is important that the underlying reason for missing data is reported, and ambiguous terms such as "lost to follow-up" and "withdrawal" are not used without providing the underlying reason. See Part A section 5.

9) Report a comparison of the characteristics of those with observed and missing data.

a) Report a comparison of the baseline characteristics of those with observed and missing data, for each outcome.

b) If participants with missing data are excluded from the analysis, report a comparison of the baseline characteristics between trial arms of those included in the analysis and a separate comparison of those excluded from the analysis. This may be provided as supplementary material.


Face Shiel
Direct Splash Protection
Note Please gently remove protective film before using

Comparison of the baseline characteristics between those with observed and missing data will enable the reader to assess if or how those with missing data differed from those with observed data. This will help determine the risk missing data pose to both internal and external validity, as well as the plausible missing data mechanism(s) (see Part B section 6).

A comparison of baseline characteristics between trial arms of those included in the analysis, and a separate comparison of baseline characteristics between trial arms of those excluded from the analysis should be reported[71]. These findings will allow readers to make a judgement about the post-randomisation risk that missing data pose to the internal validity of the findings.

If a statistical test is used to assess the difference between those with observed and missing data and between trial arms, it is important to specify that the analyses are not necessarily powered to detect such a difference and the absence of a difference does not guarantee lack of bias. It would be preferable not to rely solely on statistical tests when making these decisions, with interpretation of estimates and confidence intervals likely to be much more informative.

10) Report the primary analysis based on the primary assumption about the missing data mechanism, for all outcomes with missing data.

It is important that the primary analysis reported is the analysis based on the primary and most plausible assumption about the missing data mechanism. Often when conducting a missing data analysis, the first analysis conducted is the default method of complete case analysis. However, this may not be the most valid method to use given the assumptions about the missing data mechanism (see Part B introduction).

11) Report results of the missing data sensitivity analyses for all outcomes with missing data. As a minimum, a summary of the missing data sensitivity analyses for the primary outcome(s) should be reported in the main paper with the full results in the supplementary material.

Assessment of the sensitivity of the findings to different assumptions about the missing data mechanism, as part of a sensitivity analysis, is central to a principled approach to handling missing data (see Part B introduction). This information allows the reader to determine the risk that missing data and the choice of analysis pose to the study findings. For this reason, as a minimum, a summary of such information should be provided in the main paper, with the full analysis available as supplementary material.

Discussion section

12) Discuss the impact of missing data on the interpretation of findings, considering both internal and external validity.

When missing data occur, researchers should explicitly report their evaluation of the risk missing data posed to the study findings. In particular, the impact on internal and external validity should be reported.

Conclusion

There is an urgent need to reduce the considerable waste in clinical research and improve its value for citizens, their loved ones and society.

Missing data are ubiquitous in healthcare research and are a major issue in palliative and end of life care research in particular.

This guidance aimed to bring together and build upon the current expertise, recommendations and evidence on how to reduce, handle and report missing data in palliative and end of life care trials, in order to provide a framework on how research waste due to missing data can be addressed.

Palliative care trials encompass a diverse range of studies that have varying rates and risks posed by missing data[13]. Each trial team will have to consider carefully how best to reduce and handle missing data for their particular trial, balancing the risk of missing data against the resource implications.

These recommendations provide an overview of how missing data can be tackled and should be considered for each trial and implemented in accordance with the risk presented by missing data given the design and conduct of that particular trial.

This guidance aims to be useful and informative for all members of the multidisciplinary team involved in the design, conduct, analysis, interpretation and implementation of palliative care trials. An inclusive approach was undertaken to provide all stakeholders with a broad understanding of the importance of, and how to address, the issue of missing data. This is crucial for dealing with missing data and to improve the implementation of missing data recommendations – which to date has been poor in palliative care and other fields.

The key messages for different stakeholders are:

- **Patients, carers and PPI partners** – collection of complete data is crucial to optimise the value of research. If this is not possible, understanding the reasons for this is very important.

- **Trialists** – trial design and conduct should prioritise minimising missing data from the start.

- **Data analysts** – missing data should be handled in a principled way, based on plausible assumptions about the missing data. Missing data sensitivity analyses should assess the sensitivity of the findings to different plausible assumptions about the missing data. In palliative and end of life care trials in particular, how data truncated due to death are handled needs careful consideration.

- **Funders, publishers and users of research** – missing data and how they are handled must be reported clearly and transparently. Careful review of the amount of missing data and whether there is evidence of different amounts or types of people with missing data in each trial arm is required. If differences

between trial arms is observed, this would suggest that the missing data may have introduced bias, especially if they were not analysed appropriately. Funders should consider their role in encouraging the implementation of missing data methods guidance.

Further research is required to develop the evidence base on how to effectively reduce missing data, handle it appropriately and improve the reporting of trials. This guidance should be updated as new evidence, experience and techniques develop, and in due course it will be important to evaluate the impact of the guidance and how to improve its implementation.

izusek/iStockphoto.com

Further reading and resources

A. Reducing missing data

General strategies to reduce missing data in trials

- Panel on Handling Missing Data in Clinical Trials. The prevention and treatment of missing data in clinical trials 2010. [Online] [Accessed Jan 2020]. Available from: *nap.edu/catalog/12955/the-prevention-and-treatment-of-missing-data-in-clinical-trials*

- European Medicines Agency. European Medicines Agency guideline on missing data in confirmatory clinical trials 2010. London: EMA. [Online] [Accessed Jan 2020]. Available from: *ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.*

- Brueton V, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I and Rait G. Strategies to improve retention in randomized trials. Cochrane Database Syst Rev 2013. *bmjopen.bmj.com/content/4/2/e003821*

- Bower P, Brueton V, Gamble C, Treweek S, Tudor-Smith C, Young B and Williamson P. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. Trials 2014; 15: 399. *trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-15-399*

Strategies to reduce missing data in palliative and end of life care trials

- Hussain JA, White IR, Johnson MJ, Bland M, Currow DC. Performance status and trial site-level factors are associated with missing data in palliative care trials: An individual participant-level data analysis of 10 phase 3 trials. Palliative Medicine. 2021; 35:1998-2007. *journals.sagepub.com/doi/10.1177/02692163211040970*

- Hussain JA, White IR, Langan D, Johnson MJ, Currow DC, Torgerson DJ and Bland M. Missing data in randomized controlled trials testing palliative interventions pose a significant risk of bias and loss of power: a systematic review and meta-analyses. Journal of Clinical Epidemiology 2016; 74: 57-65. *researchgate.net/publication/287798097_Missing_data_in_randomised_controlled_trials_testing_palliative_interventions_pose_a_significant_risk_of_bias_and_loss_of_power*

- Preston NJ, Fayers P, Walters SJ, Pilling M, Grande GE, Short V, Owen-Jones E, Evans CJ, Benalia H, Higginson IJ, Todd CJ and MOREcare. Recommendations for managing missing data, attrition and response shift in palliative and end-of-life care research: part of the MORECare research methods guidance on statistical issues. Palliative Medicine 2013; 27: 899-907. DOI: 10.1177/0269216313486952.

Studies Within A Trial (SWAT)

- Treweek S, Bevan S, Bower P, Campbell M, Christie J, Clarke M, Collett C, Cotton S, Devane D, El Feky A, Flemyng E, Galvin S, Gardner H, Gillies K, Jansen J, Littleford R, Parker A, Ramsay C, Restrup L, Sullivan F, Torgerson D, Tremain L, Westmore M and Williamson PR. Trial Forge Guidance 1: what is a Study Within A Trial (SWAT)? Trials 2018; 19: 139. *trialsjournal.biomedcentral.com/articles/10.1186/s13063-018-2535-5*

B. Handling missing data

How to handle missing data (non-technical explanation)

- White IR, Horton NJ, Carpenter J and Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ 2011; 342: d40. *bmj.com/content/342/bmj.d40*

- Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP and Stern H. The prevention and treatment of missing data in clinical trials. New England Journal of Medicine 2012; 367: 1355-1360. *nejm.org/doi/full/10.1056/nejmsr1203730*

- Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM and Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009; 338: b2393. *bmj.com/content/338/bmj.b2393*

- Carpenter J and Kenwood M. Guideline for handling missing data in social science research. London School of Hygiene & Tropical Medicine. [Online] [Accessed Jan 2020]. Available from: *lshtm.ac.uk/media/37311*

- Bartlett J and Carpenter J. Module 14: Missing data – Concepts 2013. National Centre for Research Methods. [Online] [Accessed 2016 January]. Available from: *https://www.cmm.bris.ac.uk/lemma/pluginfile.php/16375/mod_resource/content/2/Module_14_LEMMA_Concept_MissingData_final.pdf*

How to analyse missing data (technical explanation)

- Carpenter JR, Kenward MG. Multiple Imputation and Its Application. Wiley; 2013. *onlinelibrary.wiley.com/doi/book/10.1002/9781119942283*

- van Buuren S. Flexible Imputation of Missing Data. 2nd ed. CRC press; 2018. *stefvanbuuren.name/fimd/*

How to handle data truncated due to death

- Kurland BF, Johnson LL, Egleston BL and Diehr PH. Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. Stat Sci 2009; 24: 211. DOI: 10.1214/09-STS293.

- Colantuoni E, Scharfstein DO, Wang C, Hashem MD, Leroux A, Needham DM and Girard TD. Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. BMJ 2018; 360: j5748. *bmj.com/content/360/bmj.j5748*

- Diehr P and Johnson LL. Accounting for missing data in end-of-life research. Journal of Palliative Medicine 2005; 8 Suppl 1: S50-57. DOI: 10.1089/jpm.2005.8.s-50.

- Hussain JA, White IR, Johnson MJ, Ekstrom M and Currow DC. Analysing data in palliative care trials. BMJ 2018; 362: k2943. DOI: 10.1136/bmj.k2943.

Auxiliary variables in multiple imputation

- Hardt J, Herke M and Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. BMC Medical Research Methodology 2012; 12: 184. *bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-184*

- Collins LM, Schafer JL and Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological Methods 2001; 6: 330-351. *researchgate.net/publication/11578765_A_Comparison_of_Restrictive_Strategies_in_Modern_Missing_Data_Procedures*

C. Reporting missing data

General criteria for reporting missing data in trials

- Moher D, Hopewell S, Schulz K, Montori V, Gotzsche P, Devereaux P, Elbourne D, Egger M and Altman D. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c869. *bmj.com/content/340/bmj.c869*

- Akl EA, Shawwa K, Kahale LA, Agoritsas T, Brignardello-Petersen R, Busse JW, Carrasco-Labra A, Ebrahim S, Johnston BC, Neumann I, Sola I, Sun X, Vandvik P, Zhang Y, Alonso-Coello P and Guyatt GH. Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide. BMJ Open 2015; 5: e008431. *bmjopen.bmj.com/content/5/12/e008431*

- Dumville JC, Torgerson DJ and Hewitt CE. Reporting attrition in randomised controlled trials. BMJ 2006; 332: 969-971. *researchgate.net/publication/7152136_Reporting_attrition_in_randomised_controlled_trials*

Reporting of missing data in palliative and end of life care trials with recommendations

- Hussain JA, Bland M, Langan D, Johnson MJ, Currow DC and 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. Journal of Clinical Epidemiology 2017; 88: 81-91. *sciencedirect.com/science/article/pii/S0895435616304735*

Developing the guidance

This guidance was developed following an evaluation of the issues missing data present to palliative care trials by Dr Jamilla Hussain as part of her NIHR-funded Doctoral Research Fellowship. This included a systematic review of trials in palliative care which assessed the extent of and factors associated with missing data in this field; an individual participant-level data analysis of ten palliative care trials, which assessed the participant and site-level factors associated with missing data; and in-depth semi-structured interviews with researchers involved in palliative care trials with a particular focus on how missing data can be reduced.

Following a discussion at the Marie Curie Strategic Advisory Committee chaired by Professor Sir Andy Haines, Marie Curie hosted a workshop on missing data in palliative care trials in October 2017. This involved PPI partners, clinicians, researchers, statisticians and methodologists. During the afternoon, delegates helped to refine and develop guidance on how best to reduce, handle and report missing data. This formed the basis of the guidelines presented in this document.

Details of the methods are available in *Development of guidelines to reduce, handle and report missing data in palliative care trials: A multi-stakeholder modified nominal group technique*[5].

Following this, a writing group comprising the authors of this guidance worked to incorporate the recommendations of the delegates and notes by workshop scribes with the expertise of the writing group. Several drafts were reviewed by the writing group and these were presented to the Marie Curie Research Strategy Advisory Committee in March 2020.

There was a substantial delay in publication as the lead author had a period of sick/maternity leave and was then redeployed to work clinically during the COVID 19 pandemic. Our aim was to provide guidance based on current practice, knowledge and evidence and we hope to update the guidance as further evidence emerges.

References

1. International council for harmonisation of technical requirements for pharmaceuticals for human use. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1), *https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf* (2019).

2. National Institute for Health Research. Pilot and feasibility studies, *https://www.journalslibrary.nihr.ac.uk/information-for-authors/report-contents/report-types/pilot-and-feasibility-studies.htm*

3. Bland JM. Missing data. *An introduction to medical statistics*. Fourth ed. Oxford: Oxford University Press, 2000.

4. Van Mechelen W, Aertgeerts B, De Ceulaer K, et al. Defining the palliative care patient: a systematic review. *Palliat Med* 2013; 27: 197-208. *researchgate.net/publication/221812349_Defining_the_palliative_care_patient_A_systematic_review*

5. Hussain JA, White IR, Johnson MJ, et al. Development of guidelines to reduce, handle and report missing data in palliative care trials: A multi-stakeholder modified nominal group technique. *Palliat Med*; 0: 02692163211065597. *journals.sagepub.com/doi/full/10.1177/02692163211065597*

6. Chalmers I and Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009; 374: 86-89. *researchgate.net/publication/40028343_Avoidable_Waste_in_the_Production_and_Reporting_of_Research_Evidence*

7. Al-Shahi Salman R, Beller E, Kagan J, et al. Increasing value and reducing waste in biomedical research regulation and management. *Lancet* 2014; 383: 176-185. *researchgate.net/publication/259697381_Increasing_value_and_reducing_waste_in_biomedical_research_regulation_and_management*

8. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014; 383: 166-175. DOI: 10.1016/S0140-6736(13)62227-8.

9. Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. *Lancet* 2014; 383: 101-104. DOI: 10.1016/S0140-6736(13)62329-6.

10. Carpenter J, Bartlett J and Kenward M. *http://missingdata.lshtm.ac.uk* (accessed December 2014).

11. McKnight PE, McKnight KM, Sidani S, Figueredo A.J. Chapter 5: Diagnostic procedures. In: Kenny DA (ed) Missing data: a gentle introduction. The Guilford Press, 2007.

12. Diehr P and Johnson LL. Accounting for missing data in end-of-life research. *Journal of Palliative Medicine* 2005; 8 Suppl 1: S50-57. DOI: 10.1089/jpm.2005.8.s-50.

13. Hussain JA, White IR, Langan D, et al. Missing data in randomized controlled trials testing palliative interventions pose a significant risk of bias and loss of power: a systematic review and meta-analyses. *J Clin Epidemiol* 2016; 74: 57-65. 2016/01/01. *researchgate.net/publication/287798097_Missing_data_in_randomised_controlled_trials_testing_palliative_interventions_pose_a_significant_risk_of_bias_and_loss_of_power*

14. 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. *bmj.com/content/344/bmj.e2809*

15. Wood AM, White IR and Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials* 2004; 1: 368-376. DOI: 10.1191/1740774504cn032oa.

16. 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. *bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-14-118*

17. Hussain JA. The reporting, extent, risk of bias and factors associated with missing data in palliative care randomised controlled trials. PhD Thesis, Hull York Medical School, Unpublished, 2018.

18. Panel on Handling Missing Data in Clinical Trials. *The prevention and treatment of missing data in clinical trials.* 2010. National Research Council.

19. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. *ICH Harmonised Tripartite Guideline: Statistical principles for clinical trials*. 1998. ICH.

20. European Medicines Agency. *European Medicines Agency guideline on missing data in confirmatory clinical trials.* 2010. London: EMA. *ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf*

21. Preston NJ, Fayers P, Walters SJ, et al. Recommendations for managing missing data, attrition and response shift in palliative and end-of-life care research: part of the MORECare research method guidance on statistical issues. *Palliat Med* 2013; 27: 899-907. DOI: 10.1177/0269216313486952.

22. Hussain JA, Bland M, Langan D, et al. 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. 2017/05/24. *sciencedirect.com/science/article/pii/S0895435616304735*

23. Hussain JA, White IR, Johnson MJ, Bland M, Currow DC. Performance status and trial site-level factors are associated with missing data in palliative care trials: An individual participant-level data analysis of 10 phase 3 trials. *Palliat Med* 2021: 2692163211040970. 2021/09/17. *journals.sagepub.com/doi/10.1177/02692163211040970*

24. Carpenter JR and Kenward MG. *Missing data in randomised controlled trials- a practical guide*. 2007.

25. Sherman DW, McSherry CB, Parkas V, et al. Recruitment and retention in a longitudinal palliative care study. Appl Nurs Res 2005; 18: 167-177. DOI: 10.1016/j.apnr.2005.04.003.

26. Terry W, Olson LG, Ravenscroft P, Wilss L, Boulton Lewis G. *Hospice patients' views on research in palliative care. Intern Med J* 2006; 36: 406-413. DOI: 10.1111/j.1445-5994.2006.01078.x.

27. Hussain JA, White IR, Johnson MJ, Bland M, Currow DC. Performance status and trial site-level factors are associated with missing data in palliative care trials: An individual participant-level data analysis of 10 phase 3 trials. *Palliat Med* 2021; 35: 1998-2007. 2021/09/17. *journals.sagepub.com/doi/10.1177/02692163211040970*

28. Brueton V, Tierney J, Stenning S, et al. Strategies to improve retention in randomised trials. *Cochrane Database Syst Rev* 2013. *bmjopen.bmj.com/content/4/2/e003821*

29. Trial Forge. What is PRIORITY II about?, *https://www.trialforge.org/priority-two/* (2018, accessed February 2019).

30. Brunsdon D, Biesty L, Brocklehurst P, et al. What are the most important unanswered research questions in trial retention? *A James Lind Alliance Priority Setting Partnership: the PRioRiTy II (Prioritising Retention in Randomised Trials) study. Trials* 2019; 20: 593. 2019/10/17. *trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3687-7*

31. Blazeby JM, Nicklin J, Brookes ST, Winstone K, Alderson D. Feasibility of quality of life assessment in patients with upper gastrointestinal tract cancer. *Br J Cancer* 2003; 89: 497-501. 2003/07/31. *nature.com/articles/6601146*

32. Moher D, Hopewell S, Schulz K, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869. *bmj.com/content/340/bmj.c869*

33. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011; 342: d40. *bmj.com/content/342/bmj.d40*

34. Dunn G, Emsley R, Liu H, et al. Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme. *Health Technol Assess* 2015; 19: 1-115, v-vi. 2015/11/13. *researchgate.net/publication/283840483_Evaluation_and_validation_of_social_and_psychological_markers_in_randomised_trials_of_complex_interventions_in_mental_health_A_methodological_research_programme*

35. Currow DC. Why don't we do more rigorous clinical research so that we can stop experimenting on patients? *J Palliat Med* 2010; 13: 636-637. DOI: 10.1089/jpm.2010.9824.

36. Pickard AS and Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. *Med Care* 2005; 43: 493-499. 2005/04/20. DOI: 10.1097/01.mlr.0000160419.27642.a8.

37. Jackson D, White IR and Leese M. How much can we learn about missing data?: an exploration of a clinical trial in psychiatry. *J R Stat Soc Ser A Stat Soc* 2010; 173: 593-612. 2010/08/17. *researchgate.net/publication/45651150_How_much_can_we_learn_about_missing_data_An_exploration_of_a_clinical_trial_in_psychiatry*

38. Treweek S, Bevan S, Bower P, et al. Trial Forge Guidance 1: what is a Study Within A Trial (SWAT)? *Trials* 2018; 19: 139. 2018/02/25. *trialsjournal.biomedcentral.com/articles/10.1186/s13063-018-2535-5*

39. NIHR. Why the NIHR's new funding stream for 'Studies Within A Trial (SWATs) is potentially game-changing, *www.nihr.ac.uk/blog/why-the-nihrs-new-funding-stream-for-studies-within-a-trial-swats-is-potentially-game-changing/11101* (2018, accessed February 2019).

40. Health Research Board. Definitive interventions and feasibility awards (2020, accessed 7th August 2020).

41. Bower P, Brueton V, Gamble C, et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. *Trials* 2014; 15: 399. 2014/10/18. *trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-15-399*

42. Kearney A, Rosala-Hallas A, Bacon N, et al. Reducing attrition within clinical trials: The communication of retention and withdrawal within patient information leaflets. *PLoS One* 2018; 13: e0204886. 2018/11/01. *journals.plos.org/plosone/article?id=10.1371/journal.pone.0204886*

43. International Conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH Harmonised Guideline: Integrated Addendum to ICH E6(R1): *Guideline for Good Clinical Practice E6 (R2)*. 1996. Online: ICH.

44. Dunn G, Maracy M and Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Stat Methods Med Res* 2005; 14: 369-395. 2005/09/24. DOI: 10.1191/0962280205sm403oa.

45. Little RJA and Rubin DB. Statistical analysis with missing data. Second ed. *New York: Wiley,* 2002.

46. Allison PD. Missing data. Sage University papers series on quantitative applications in the social sciences. Thousand Oaks, *CA: Sage,* 2001.

47. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; 367: 1355-1360. *nejm.org/doi/full/10.1056/nejmsr1203730*

48. Bartlett J and Carpenter J. Module 14: Missing data - Concepts, *https://www.cmm.bris.ac.uk/lemma/pluginfile.php/16375/mod_resource/content/2/Module_14_LEMMA_Concept_MissingData_final.pdf* (2013, accessed January 2016).

49. White IR and Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med* 2005; 24: 993-1007. 2004/12/01. DOI: 10.1002/sim.1981.

50. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338: b2393. *bmj.com/content/338/bmj.b2393*

51. McKnight PE, McKnight KM, Sidani S, K.M, Sidani, S, Figueredo A.J. Missing data: A Gentle Introduction. *New York: The Guilford Press*, 2007.

52. Pigott TD. A review of methods for missing data. *Educational research and evaluation* 2001; 7: 353-383.

53. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res* 2018; 27: 2610-2626. 2016/12/31. *journals.sagepub.com/doi/10.1177/0962280216683570*

54. Carpenter J and Kenwood M. Guideline for handling missing data in social science research. *lshtm.ac.uk/media/37311*

55. Kurland BF, Johnson LL, Egleston BL, Diehr PH. Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. *Stat Sci* 2009; 24: 211. DOI: 10.1214/09-STS293.

56. Rubin DB. Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death. *Stat Sci* 2006; 21: 299-309.

57. Colantuoni E, Scharfstein DO, Wang C, et al. Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. *BMJ* 2018; 360: j5748. 2018/01/05. *bmj.com/content/360/bmj.j5748*

58. Hussain JA, White IR, Johnson MJ, Ekstrom M, Currow DC. Analysing data in palliative care trials. *BMJ* 2018; 362: k2943. 2018/07/12. DOI: 10.1136/bmj.k2943.

59. van Andel G, Bottomley A, Fossa SD, et al. An international field study of the EORTC QLQ-PR25: a questionnaire for assessing the health-related quality of life of patients with prostate cancer. *Eur J Cancer* 2008; 44: 2418-2424. *researchgate.net/publication/23240766_An_international_field_study_of_the_EORTC_QLQ-PR25_A_Questionnaire_for_assessing_the_health-related_quality_of_life_of_patients_with_prostate_cancer*

60. Sylvestre Y. CONSORT: missing missing data guidelines, the effects on HTA monograph reporting. *Trials* 2011; 12: A61. *trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-12-S1-A61*

61. Hardt J, Herke M and Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. *BMC Med Res Methodol* 2012; 12: 184. *bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-184*

62. Hui D, Glitza I, Chisholm G, Yennu S, Bruera E. Attrition rates, reasons, and predictive factors in supportive care and palliative oncology clinical trials. *Cancer* 2013; 119: 1098-1105. *acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.27854*

63. Hussain J. Why we need to tackle the problem of missing data in palliative care research. Online: *Marie Curie*, 2017. *mariecurie.org.uk/blog/why-we-need-to-tackle-missing-data/170450*

64. Harrison P, Pointon JJ, Chapman K, Roddam A, Wordsworth BP. Interleukin-1 promoter region polymorphism role in rheumatoid arthritis: a meta-analysis of IL-1B-511A/G variant reveals association with rheumatoid arthritis. *Rheumatology (Oxford)* 2008; 47: 1768-1770. *researchgate.net/publication/23302846_Interleukin-1_promoter_region_polymorphism_role_in_rheumatoid_arthritis_A_meta-analysis_of_IL-1B-511AG_variant_reveals_association_with_rheumatoid_arthritis*

65. Li T, Hutfless S, Scharfstein DO, et al. Standards should be applied in the prevention and handling of missing data for patient-centered outcomes research: a systematic review and expert consensus. J *Clin Epidemiol* 2014; 67: 15-32. DOI: 10.1016/j.jclinepi.2013.08.013.

66. Kistin CJ. Transparent reporting of missing outcome data in clinical trials: applying the general principles of CONSORT 2010. *Evid Based Med* 2014; 19: 161-162. *ebm.bmj.com/content/19/5/161*

67. 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. *bmjopen.bmj.com/content/5/12/e008431*

68. Calvert M, Blazeby J, Altman D, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013; 309: 814 - 822. *jamanetwork.com/journals/jama/fullarticle/1656259*

69. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004; 141: 781-788. *acpjournals.org/doi/10.7326/0003-4819-141-10-200411160-00009*

70. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; 345: e5661. *bmj.com/content/345/bmj.e5661*

71. Dumville JC, Torgerson DJ and Hewitt CE. Reporting attrition in randomised controlled trials. *BMJ* 2006; 332: 969-971. *researchgate.net/publication/7152136_Reporting_attrition_in_randomised_controlled_trials*

Appendix 1

Membership of the Marie Curie Research Strategic Advisory Committee (Between 2017 and 2019)

Member	Affiliation
Professor Sir Andy Haines (Chair)	London School of Hygiene & Tropical Medicine
Professor Tim Peters (Deputy Chair)	Professor Tim J Peters, Bristol Medical School and Bristol Dental School, University of Bristol
Professor Declan Walsh	Levine Cancer Institute, United States, and Marie Curie Trustee
Professor Ian Tannock	Princess Margaret Cancer Centre, Toronto, Canada
Professor John Norrie	University of Edinburgh
Professor Chris Ecclestone	University of Bath
Professor Phil Hannaford	University of Aberdeen
Professor Jane Seymour	University of Sheffield
Professor Bee Wee	Oxford University Hospitals NHS Foundation Trust
Professor Allan Kellehear	University of Bradford
Professor Mike Bennett	University of Leeds
Professor Sue Latter	University of Southampton
Professor Joanna Coast	University of Bristol
Dr Kathy Seddon	Marie Curie Research Expert Voice
Peter Buckle	Marie Curie Research Expert Voice

Appendix 2

Blog by Dr Jamilla Hussain regarding the challenges of missing data in palliative and end of life care research.

The blog can also be accessed on Marie Curie's website.

Why we need to tackle the problem of missing data in palliative care research

Nearly nine out of every ten randomised controlled trials in palliative care report having missing data, which can mean these studies are biased, less accurate and less powerful than they should be.

It's vital that we find ways to reduce the amount of data missing in these trials, and make the best use of the data we do have available. This is so we can make sure studies produce valid results that help reduce the uncertainty around how palliative care can help others in the future.

1) Why is missing data a problem in palliative care?

When a study sets out to gather information, but for whatever reason it's never collected, this is known as "missing data".

As a palliative care doctor, my patients tell me they want to take part in research in order to help others. These remarkable individuals donate their limited time and energy in order to give something back through research - even when this means spending less time with friends and family, or simply resting.

Sadly, some people become too poorly to finish a study or are unable to provide all the information needed for other reasons, such as moving home.

On average, just under a quarter of the information we need at the most important point in a study (known as the primary end-point) is missing in palliative care trials.

This is over twice as much as that found in other areas of healthcare.

For more information on this, see the full research, *Missing data in randomized controlled trials testing palliative interventions pose a significant risk of bias and loss of power: a systematic review and meta-analyses.*

2) Why having all the data matters for valid results in medical trials

Missing data is common in randomised controlled trials – studies that investigate whether one treatment (e.g. treatment A) is better than another treatment (e.g. treatment B).

To carry out these trials, we have to make sure the people receiving treatment A are similar to those receiving treatment B.

This is so that if one group does better than the other group, we can say this is likely to be due to the treatment and not because of other differences between the people who were in the groups – see this article in Healthtalk about understanding allocation.

How a randomised controlled trial works

In a randomised controlled trial, we make sure the two groups are similar by allocating participants at random to a treatment group. You can see the steps in a randomised controlled trial in figure 1.

Why missing data causes biased results

Missing data can affect the balance of the groups in a trial so they are no longer similar. This means we are less certain that any differences we find between those who get treatment A and those who get treatment B are due to the treatment alone.

This can happen, for instance, if each group has a different number of people with missing data.

For example (figure 2), if two pears and one apple have missing data in Group A and only one pear and one apple in Group B, we are no longer comparing groups made up of similar numbers of apples and pears.

The grey areas show which people have missing data, and how this can affect the balance of the groups.

In this scenario, because most of the data from Group A is from apples, we might mistakenly think that treatment A caused most of the fruit to shrink.

Results can also be biased if data from different types of people are missing from each group (figure 3). For example, if all the apples have missing data in Group A and all the pears have missing data in Group B.

If the data from different types of people is missing from each group, we will no longer be comparing similar groups of people.

Figure 1

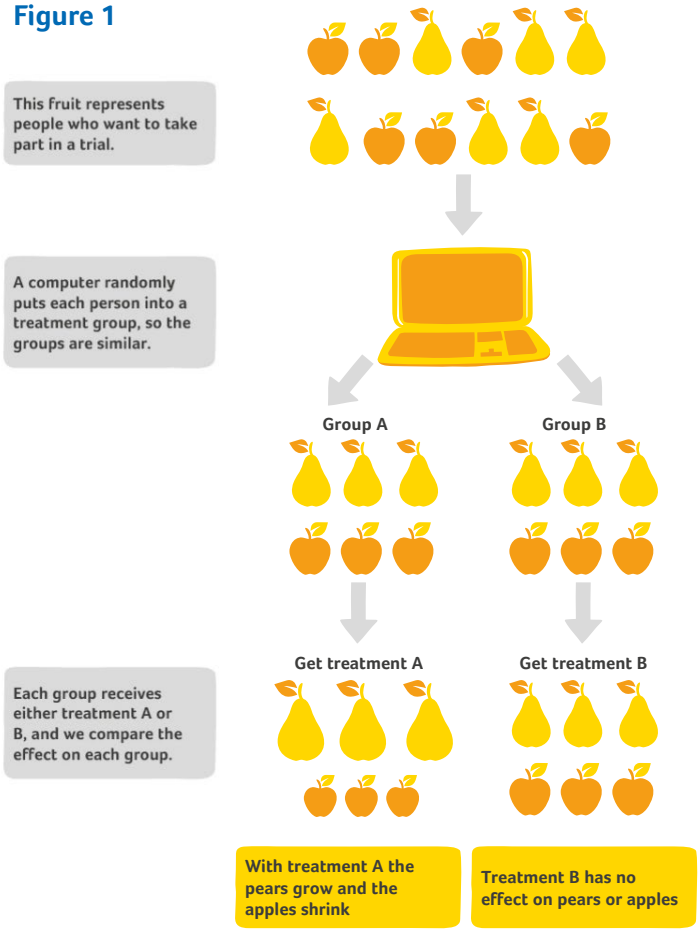


Figure 2

Figure 3

Even though this time we still have the same number of fruit overall in each group (i.e. three in each group), we are now comparing apples with pears.

So we might mistakenly say that treatment A helps all fruit to grow. But this would be wrong, because no data from apples are available in that group.

3) Missing data reduces the accuracy and power of a study

Before trials start, researchers calculate the minimum number of people they need to take part in the study. The more people there are in a study, the more precise the findings will be, and the more powerful the study is to find the effect the treatment has if one exists.

Studies usually try not to have more people than is necessary. Most also estimate how many people will have missing data and increase the minimum number needed to take account of this.

If there is more missing data than expected however, then the study will not have enough power (i.e. people) to tell whether, for example, treatment A is better than treatment B. So the study cannot do what it set out to do.

4) Failing to help those who need it most

In palliative care research, a common reason for missing data is that the person becomes too unwell to continue to take part. This means the information on how the treatment affects the people who are the most poorly is often missing.

This is a big problem for palliative care teams, because we are responsible for taking care of the people who become too poorly and need to make decisions about how best to treat them.

So if we don't know if the treatment works for these people, we have to rely on information from people who are not as frail and unwell – and it's difficult to know if this will result in benefit or harm.

5) Incomplete data is wasted data

In six out of ten palliative care trials, *researchers don't include any of the data from the people who didn't provide a complete set*.

This means the information from those people who could only provide part of the data isn't included in the results.

For the study, this is wasteful and will often bias the results. But more importantly, this means for those people volunteering their time to take part in research, but who can only provide some of the data, their time and energy is also wasted.

Fortunately, there are ways researchers can make sure this data is analysed. But we now need to find ways to get everyone to use them.

6) What next?

There is no complete cure for missing data, but it's vital we identify ways to reduce the amount of missing data wherever possible and make the best use of the data we have.

Marie Curie recently hosted a workshop for researchers, clinicians and other partners in research on this subject. We discussed how to reduce missing data, what to do with missing data when analysing results of a study, and how to report missing data clearly so we can understand the impact.

To find out more or learn about the outcomes, contact Jamilla Hussain at jamilla.hussain@bthft.nhs.uk or follow @JamillaHussain1 on Twitter.

For more information:

Dr Jamilla Hussain
E: jamilla.hussain@bthft.nhs.uk