CONTINUOUS MISSING PARTICIPANT DATA IN RANDOMIZED CONTROLLED TRIALS

REPORTING, EXTENT, AND STATISTICAL APPROACHES TO DEAL WITH CONTINUOUS MISSING PARTICIPANT DATA IN RANDOMIZED CONTROLLED TRIALS

By YUQING ZHANG, M.D., M.Sc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2015)

Hamilton, Ontario (Health Research Methodology)

TITLE: Reporting, extent, and statistical approaches to deal with continuous missing participant data in randomized controlled trials

AUTHOR: Yuqing Zhang, M.D. (Beijing University of Chinese Medicine) M.Sc. (Beijing University of Chinese Medicine)

SUPERVISOR: Distinguished Professor Gordon Guyatt

NUMBER OF PAGES: xv, 128

ABSTRACT

Background and Objectives:

Missing participant data are likely to bias the results of randomized control trials (RCTs) when the reason for missingness is associated with status on the outcome of interest. Unlike dichotomous MPD in RCTs, which have been thoroughly investigated, knowledge regarding continuous MPD in RCTs is much more limited. Our objectives were 1) using an adapted checklist, to assess the reporting quality of simulation studies comparing methods to deal with continuous MPD; 2) identify optimal methods proposed by biostatisticians and tested in simulations studies for continuous MPD in RCTs; 3) evaluate how authors report MPD, and how they plan and conduct analyses to deal with MPD in RCTs.

Methods:

We conducted two systematic surveys. The first identified methods papers published till 2015 January that compared statistical approaches to deal with continuous MPD in RCTs using at least one simulation. In this sample, we considered both the quality of reporting and the results. The second survey identified a representative sample of individual RCTs published in 2014 in core journals reporting the results of at least one continuous variable addressing a patient-important outcome.

Results and conclusion:

Our survey identified important limitations in reporting quality of simulation studies that compared statistical approaches to deal with continuous MPD, particularly in the reporting of simulation procedures. Only one of 60 studies reported the random number generator used and none reported starting seeds or failures during simulation. Less then half reported software used to perform simulation (41.7%) or analysis (48.3%), and only 4 (5%) reported justification of number of simulations. When facing continuous MPD in RCTs, results of simulation studies demonstrate that trialists seeking optimal approaches may choose robust regression or mixed models and avoid using last observation caring forward. Continuous MPD frequently occurs in RCTs and the extent is typically substantial (median greater than 10%). Methods sections in trial reports typically do not provide adequate detail on how they dealt with MPD in their primary analysis. Among methods actually implemented to deal with MPD, most authors use only available data, thus excluding MPD from the analysis. Seldom do investigators apply statistical approaches to impute or taking into account of MPD nor conduct sensitivity analysis to address the impact of it.

A comprehensive knowledge synthesis summarizing current available statistical approaches and its relative merits, as well as the current used methods in RCTs provide clear implications on how the practise of using methods to handle continuous MPD should shift in individual RCTs. Trialists should use mixed models and robust regressions and avoid using last observation caring forward method.

Preface

This thesis has been conducted as a "sandwich thesis" and consists of three individual manuscripts/papers submitted to journals for publications. These are:

1.) Chapter 1: Introduction of the thesis
2.) Chapter 2: Reporting quality of simulation studies comparing statistical methods of handling MPD for individual RCTs: A systematic survey; manuscript
3.) Chapter 3: Handling trial participants with missing outcome data for continuous outcomes in randomized control trials: a systematic survey of the methods literature; manuscript
4.) Chapter 4: Reporting and analysis of missing participant continuous data in randomized controlled trials: a systematic survey; manuscript
5.) Chapter 5: Discussion and conclusion/summary

At the time of writing (June 2015-July 30th 2015) all three chapters have been written and submitted for publication.

Copyright permission:

Acknowledgments:

The thesis work would have been impossible without enormous physical and spiritually support and help from many individuals along the way. I would like to thank them and express my deepest appreciation.

I would love to express my most special thanks to Dr. Gordon Guyatt for his incredible support, guidance, supervision and mentorship along the way. He is best role model for my life in so many aspects. I deeply appreciate the opportunity to work with him and all the great things I have learnt from him in and out of work. I would also thank my thesis committee, Dr Lehana Thabane, Dr Holger J Schünemann, and Dr. Elie Akl for your guidance, support, and inspiration for my thesis work and for my career.

I have to thank the team of friends and colleagues who worked with me along the way for all of their extremely hard work, effort, dedication, and the time they invested in these work.

My family has provided enormous amount of support along the way for my whole education till now.

Lastly, I would like to thank the colleagues, friends, and staff in the department of Clinical Epidemiology and Biostatistics for all the academic and non-academic support to make this challenging journey more enjoyable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIEXS

List of Abbreviations

MPD: Missing participant data

RCT: Randomized controlled trials

MCAR: Missing completely at random

MAR: Missing at random

NMAR: Not missing at random

LOCF: Last observation caring forward

GEE: Generalized estimating equation

MI: Multiple imputation

SR: Systematic review

Declaration of Academic Achievement

Dr. Zhang was the principle contributor and first author of all the manuscripts contained in this thesis. Dr. Zhang played the primary role in the conception and design of the study, with the aid of her supervisor Dr. Guyatt and committee members. Y. Zhang made following contribution in all papers included in the dissertation: developing the research questions, writing protocols and analysis plans, designing search strategies, designing screening and data abstraction forms, pilot testing, calibration and management the team of co-authors, conducting statistical analyses, designing figures and tables, writing all manuscripts, submitting the manuscripts to her supervisor and co-authors and addressing feedbacks. The co-authors contributed in selecting, acquiring and analysing the data as well as providing input for manuscripts in preparation for submission for publication. All studies were conducted between summer 2014 and summer 2015. Details of the contributions of coauthors are described at the end of each manuscript.

CHAPTER 1:

INTRODUCTION

Missing Participant Data

Missing participant data (MPD) refers to outcome data from trial participants that are not available to include in the data analysis. MPD does not include missing studies (unpublished) or unreported outcomes that planned in the protocol but not reported in the trials[1]. A broadly used classification on reasons of missing data, which also refers to missing mechanisms, includes the following three classes[2,3]. Missing completely at random (MCAR) indicates the reason for missing observations is unrelated to either observed or unobserved outcome or covariates. For instance, patients missed one follow-up due to a car accident. In this case, ignoring missing data and only analyzing the available data will not cause bias but only reduce power. MCAR is a strong assumption and almost always unlikely in trials[4].

Missing at random (MAR), a more reasonable assumption, occurs when the missing observations are related to observed outcomes or covariates (i.e. patient characteristics). For instance, older patients. In this case, ignoring MPD will cause bias if the patients with missing data differ in age in intervention and control groups. When data are MAR statistical approaches need to be applied to reduce bias.

Lastly, and likely the most realistic assumption for at least some of the missing data in randomized trials is missing not at random (NMAR). When data is NMAR, the missing

observations are associated with unobserved data even after considering apparent associations in the observed data[5,6]. Dealing with NMAR requires conducting sensitivity analyses to test the robustness of the primary analysis under various reasonable assumptions about the nature of he missing data[3]. Trial investigators are rarely able to provide confident judgements about whether MPD is MCAR or MAR, NMAR. Therefore, more conservative approache in which one tries to make inferences about missing data on the basis of the association between patient characteristics and outcomes, and in addition one conducts sensitivity analyses, is advisable for continuous outcomes in randomized trials.

MPD in RCTs

MPD in randomized controlled trial (RCT) is commonly used interchangeably with loss to follow-up, discontinued prematurely, or outcome not assessable[1]. RCT is a widely used optimal type of study design that provides most trustworthy evidence regarding the efficacy of interventions[7,8]. MPD in RCT threaten the prognostic balance between intervention and control groups and may bias the result. For example, for the outcome of symptoms, if patients in intervention group with more severe symptoms withdraw more frequently than similar patients in the control group, the result will be an overestimate due to the MPD.

MPD may therefore bias estimates of intervention effects. MPD, when omitted, might also lead to a reduction in statistical power.

Although trial investigators applied various strategies and significant endeavours to minimize MPD, its presence is in most case unavoidable[9]. Investigators have reported MPD in up to 89% of RCTs across all therapeutic areas[10-13]. Moreover Akl et al[13] found up to one third of RCTs published in leading medical journals may lose statistical significant result when applying plausible assumptions regarding the missing data. It is therefore crucial to appropriately deal with MPD in RCTs to minimize biased results.

Statistical approaches to handle MPD

Much attention has been given in the past decades to missing data in RCTs. Biostatisticians suggested various methods to deal with MPD in RCTs[14-18]. The statistical approaches can be classified in four broad categories: data deletion, single imputation, multiple imputation, and data augmentation[19].

One of the most frequently seen and straightforward approaches within data deletion is complete case analysis, which refers to omitting the patients with one or more missing values[20]. It is easy to apply in analysis and may yield unbiased result when data are MCAR but only reduces power[2,21,22]. The other data deletion approach is all available data analysis also called available case method, which refers to using all available observations even patients have some MPD through the entire follow-up when outcomes are measured repeatedly. For instance, if one patient missed three follow-ups out of total ten follow-ups, all available data analysis would still include the remaining seven observations in the analysis[19].

The basic principle for imputation methods is to substitute each missing observation with a new one and then conduct the analysis as if no MPD were present. Single imputation

replaces all missing observations with a single value and it therefore reduces the true variability in the data and artificially narrows the confidence interval[23]. Mean imputation, hot-deck, cold-deck, and last observation carried forward (LOCF) are commonly seen single imputation methods[19].

In the contrast, multiple imputations produce multiple values and replaces the missing observations with more than one plausible value to complete multiple alternative complete datasets[2]. Then each dataset is analyzed independently to obtain the effect estimate. Lastly, multiple parameter estimates are combined and obtain one single best estimate[17]. Multiple imputation takes into account the variability of missing values and incorporates uncertainty [2,21,24].

Data augmentation refers to a method to invoke an algorithm to account for observed and missing data, the relationship between these two, as well as underlying statistical assumptions to estimate parameters of interest. It does not involve replacement of missing values[19]. Model based approaches belong to data augmentation. Mixed effects models, robust regression, generalized estimating equation (GEE) are frequently investigated categories and they are based on maximum likelihood[25,26], pseudo-likelihood or maximum inference [27,28], and quasi-likelihood inferences respectively[29,30].

Simulation studies and MPD

Simulation studies use computational procedures to test hypotheses and assess the performance of statistical methods in different scenarios in relation to truth[31]. Simulation studies have been widely used in the medical literature[32]. High quality simulation studies present compelling evidence on the relative merits of statistical methods in complex

situations for data from RCTs and other study designs. However, readers of simulations face the challenge of understanding the process of the simulation studies, interpreting conflicting results presented in various formats, assessing the validity of result, critically appraising the conclusion and making inferences from it. Unsatisfactory or non-uniform reporting hinders the appraisal process and may lead to misleading interpretation[31,33].

Objectives and outlines

The general objective of this thesis is to *assess the reporting quality of simulation studies on methods of handling continuous MPD, to identify optimal methods for dealing with continuous MPD in simulations, and to investigate the reporting, extent of MPD, planned and conducted analysis in RCTs.*

The thesis is a "sandwich" composed of three papers (chapters 2-4) exploring issues around continuous MPD in RCTs.

Chapter 2 contains a modified checklist for reporting of simulation studies based on Burton et al[31] and a systematic survey on the adherence of simulation studies focusing on proposing statistical approaches to deal with continuous MPD in parallel design RCTs using the modified checklist.

Having first assessed the quality of reporting in the simulations studies, we further investigate the relative merits of the compared methods in simulations. Chapter 3 is a systematic survey on all available statistical methods to deal with continuous MPD in RCTs that are empirically tested in simulations. We also compare and present the optimal and inferior methods based on their relative performance regarding bias, precision, accuracy, type I error, coverage, power and overall performance based on above criteria.

Having established the performance of existing statistical approaches to deal with continuous MPD, we sought to investigate how RCT investigators implement methods in planning and conducting analysis to deal with MPD. Chapter 4 assesses how RCT investigators report and the extent of MPD for continuous outcomes, and the analytical approaches planned and conducted in their primary and sensitivity analysis addressing MPD.

The whole thesis work aims to expound a compelling body of evidence to inform trialists, methodologists, biostatisticians, journal editors and users of the medical literatures of the reporting and methods to deal with continuous MPD in RCTs.

Overlap in material covered:

We design chapter 2 and 3 as independent manuscripts but abstract different sets of information from the same studies. Therefore, there are overlaps in the eligibility criteria, search strategy, study selection and in the descriptions of general study characterises in chapter 2 and 3.

Research context:

The series of works focus on continuous MPD in RCTs and are related to a series of parallel investigations on dichotomous MPD in RCTs and systematic review (SR)s from our research group. Further inspection of these series of studies may facilitate the understanding of MPD from a general perspective. Prior to this work, our group addressed the reporting, analytical approaches, and the impact of dichotomous MPD in five leading general medical journals (LOST-IT)[13]. Other ongoing investigations include the exploration of impact of MPD for dichotomous outcomes on pooled effect estimates

in SRs[1], and a systematic assessment on approaches of dealing with dichotomous MPD in SR (submitted). Finally, chapter 1 and 2 inspired the design, conduct and analysis of the other 2 ongoing projects on reporting and performance of statistical approaches on dichotomous outcomes in RCTs.

Reference:

1. Akl EA, Kahale LA, Agarwal A, et al. Impact of missing participant data for dichotomous outcomes on pooled effect estimates in systematic reviews: a protocol for a methodological study. *Syst Rev.* 2014;3:137.
2. Little RJ, Rubin DB. *Statistical analysis with missing data.* New York: Wiley; 2002.
3. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med.* 2013;86(3):343-358.
4. Fiero M, Huang S, Bell ML. Statistical analysis and handling of missing data in cluster randomised trials: protocol for a systematic review. *BMJ Open.* 2015;5(5):e007378.
5. 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.
6. DB. R. Inference and missing data. *Biometrika* 1976;63:581–592.
7. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ.* 2001;323(7303):42-46.
8. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg.* 2012;10(1):28-55.
9. Scharfstein DO, Hogan J, Herman A. On the prevention and analysis of missing data in randomized clinical trials: the state of the art. *J Bone Joint Surg Am.* 2012;94 Suppl 1:80-84.
10. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials.* 2004;1(4):368-376.
11. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ.* 1999;319(7211):670-674.
12. Baron G, Boutron I, Giraudeau B, Ravaud P. Violation of the intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. *Arthritis Rheum.* 2005;52(6):1858-1865.
13. 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.
14. Demirtas H. Bayesian analysis of hierarchical pattern-mixture models for clinical trials data with attrition and comparisons to commonly used ad-hoc and model-based approaches. *J Biopharm Stat.* 2005;15(3):383-402.
15. Demirtas H. Multiple imputation under Bayesianly smoothed pattern-mixture models for non-ignorable drop-out. *Stat Med.* 2005;24(15):2345-2363.
16. DeSouza CM, Legedza AT, Sankoh AJ. An overview of practical approaches for handling missing data in clinical trials. *J Biopharm Stat.* 2009;19(6):1055-1073.
17. G. L. Gadbury CSCaDBA. Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF. *Obes Reb.* 2003;4(3):175-184
.

18. Groenwold RH, Donders AR, Roes KC, Harrell FE, Jr., Moons KG. Dealing with missing outcome data in randomized trials and observational studies. *Am J Epidemiol.* 2012;175(3):210-217.
19. McKnight PE MK SS, Figueredo AJ. *Missing Data: A Gentle Introduction.* New York, NY: Guilford; 2007.
20. Allison PD. *Missing data.* Thousand Oaks,CA: Sage; 2002.
21. Little RJA RD. *Statistical analysis with missing data.* New York: Wiley; 2002.
22. Molenberghs G, Kenward MG. . *Missing data in clinical studies.* New York: Wiley; 2007.
23. Molnar FJ, Hutton B, Fergusson D. Does analysis using "last observation carried forward" introduce bias in dementia research? *CMAJ.* 2008;179(8):751-753.
24. Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *J Clin Epidemiol.* 2013;66(9):1022-1028.
25. Siddiqui O. MMRM versus MI in dealing with missing data--a comparison based on 25 NDA data sets. *J Biopharm Stat.* 2011;21(3):423-436.
26. Tsonaka R, Verbeke G, Lesaffre E. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics.* 2009;65(1):81-87.
27. Liang H, Su H, Zou G. Confidence Intervals for A Common Mean with Missing Data with Applications in AIDS Study. *Comput Stat Data Anal.* 2008;53(2):546-553.
28. LIN DY. Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics* 2003;4(3):385–398.
29. James M. Robins AR, Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association.* 1995;90:106-121.
30. Andrea Rotnitzky JMRaDOS. Semiparametric Regression for Repeated Outcomes with Nonignorable Nonresponse. *Journal of the American Statistical Association.* Dec.1998;93(444):1321-1339.
31. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med.* 2006;25(24):4279-4292.
32. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol.* 2004;19(8):751-760.
33. Samaan Z, Mbuagbaw L, Kosa D, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *J Multidiscip Healthc.* 2013;6:169-188.

CHAPTER 2:

REPORTING QUALITY OF SIMULATION STUDIES COMPARING STATISTICAL METHODS FOR HANDLING MISSING PARTICIPANT DATA IN INDIVIDUAL RANDOMIZED CONTROLLED TRIALS: A SYSTEMATIC SURVEY

Yuqing Zhang[1], Akram Alyass[1], Thuva Vanniyasingam[1], Behnam Sadeghirad[1,3], Iván D. Flórez, [1,4] Sathish Chandra Pichika[1], Sean Alexander Kennedy[5], Ulviya Abdulkarimova[6], Yuan Zhang[1], Tzvia Iljon[6], Gian Paolo Morgano[1], Luis E. Colunga Lozano[7], Fazila Abu Bakar Aloweni[8], Luciane C Lopes[9,10], Juan José Yepes-Nuñez[1,4], Yutong Fei[11], Li Wang[12], Lara A Kahale[13], David Meyre[1,2], Elie A Akl[14], Lehana Thabane[1], Gordon Guyatt [15]

[1] Department of Clinical Epidemiology and Biostatistics, McMaster University, Canada
[2] Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada
[3] Regional Knowledge Hub and WHO Collaborating Centre for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[4] University of Antioquia, School of Medicine, Medellín, Colombia
[5] Department of Diagnostic Radiology, University of Toronto, Canada
[6] Department of Mathematics and Statistics, McMaster University, Canada
[7] Hospital Civil de Guadalajara, "Fray Antonio Alcalde", México
[8] Singapore General Hospital, Singapore
[9] Universidade de Sorocaba, São Paulo, Brasil

[10] Universidade Estadual Paulista “Julio de Mesquita Filho”São Paulo, Brazil
[11] Beijing University of Chinese Medicine, McMaster University, China, Canada
[12] Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Canada
[13] Department of Internal Medicine, American University of Beirut, Lebanon
[14] Clinical Epidemiology Unit and Center for Systematic Reviews in Health Policy and Systems Research (SPARK), American University of Beirut, Lebanon
[15] Department of Medicine and Clinical Epidemiology and Biostatistics, McMaster University, Canada
Corresponding authors: Gordon Guyatt
E-mail of Guyatt: guyatt@mcmaster.ca

Abstract

Background: Addressing missing participant data (MPD) presents a serious challenge in randomized controlled trials (RCTs) focusing on continuous outcomes. Simulation studies provide a valuable method of assessing the performance of statistical methods addressing MPD. The reporting quality of simulation studies addressing MPD for individual RCTs has not been addressed thus far.

Objective: To adapt an existing checklist for simulation studies to deal with quality of reporting and to conduct a systematic survey of the reporting quality of simulation studies of methods used to handle missing data in RCTs.

Method: We conducted a systematic search in Medline, Cochrane Library, Web of Science, Journal Storage (JSTOR) from inception to January 2015 for simulation studies testing statistical strategies for dealing with MPD focusing on continuous outcomes in RCTs. We adapted previously developed criteria for reporting quality and applied them to eligible studies.

Result: Of 16,446 identified citations, 60 studies proved eligible. Studies generally had important limitations in reporting quality, particularly in reporting simulation procedures. Less then half reported software used to perform simulations (25/60: 41.7% with 95% confidence interval of [29.2%, 54.2%]) or analysis (29/60: 48.3% [35.7%, 61.0%]), and 3 (5% [0%, 10.5%]) reported justification of the number of simulations. All studies reported scenarios evaluated, statistical methods and the criteria to evaluate their performance. Approximately 95% [89.5%, 100%] (57/60) reported the number of

simulations.

Conclusion: Current reporting of simulation studies addressing methods to deal with MPD suffered from serious limitations. Authors of simulation studies need to attend more carefully to transparently reporting all relevant aspects of their methods.

Background

Missing participant data (MPD) is common in randomized control trials (RCTs), and may seriously undermine the validity of research findings.[1] A systematic survey of RCTs reporting statistically significant results published in five prestigious general medical journals found that 87% of reported RCTs suffered from MPD in their primary outcomes[2]. Furthermore, in sensitivity analysis under plausible assumptions, up to one third of the trials lost significance[2].

In the past decades, investigators have proposed a number of statistical methods to deal with MPD. Data deletion methods, data augmentation procedures, single imputation and multiple imputations represent the four broad categories of available statistical methods[3]. Statisticians have proposed various approaches for all these methods, and applied them in real clinical trials[4] or in simulations[5]. Demonstrating the use of statistical methods in RCT data can reproduce real life scenarios. Simulation studies have, however, the advantage of assessing the performance of statistical methods in relation to the known truth and can therefore provide compelling evidence regarding the relative merits of alternative approaches[6].

High-quality simulation studies can reflect complex situations in RCTs or other study designs that closely reflect real-life data. Despite their theoretical merits, readers of simulation studies face challenges in assessing the integrity of their study designs, understanding the process of simulation, interpreting the results, and making inferences. Insufficient details in reporting may seriously hamper these assessments[6].

Reporting criteria summarized in a checklist would aid in evaluation and provide guidance to investigators in reporting their simulation studies. The reporting guidelines suggested by the Enhancing the Quality and Transparency of Health Research (EQUATOR) Network do not include a checklist for reporting of simulation studies. However, Burton *et al* have proposed a comprehensive checklist of generic issues that need to be considered when designing, conducting and reporting simulation studies[6]. Adherence to these criteria would provide transparency and thus facilitate the reproducibility and assessment of the credibility of simulation studies.

Investigators have not thus far addressed the reporting quality of simulation studies evaluating methods of dealing with MPD for individual RCTs. The aims of this article are to 1) propose a checklist for reporting of simulation studies modified from the criteria of Burton *et al*[6] and 2) conduct a systematic survey using the modified criteria to address the reporting quality of simulation studies of methods of handling missing data for individual RCTs.

Methods

Definition

MPD represents outcome data from trial participants that are not available to investigators to include in their data analysis. MPD does not refer to either: 1) missing (e.g. unpublished) studies; 2) unreported outcomes (e.g. outcomes planned to report in the protocol but excluded in the trial report.

Checklist for reporting quality of simulation studies

Using Burton *et al*'s checklist[6] that focuses on the design, conduct and reporting of simulation studies, we retained and modified items relevant to reporting. To improve replicability of studies, we added "reported software to perform analysis" to item 2. To appropriately evaluate methods based on various statistical criteria, we added precision, type I error and power to item 8. The final adapted checklist we used included:

1. Defined the aims of the simulation
2. Simulation procedures:
    - Reported dependence of simulated datasets
    - Reported starting seeds
    - Reported random number generator
    - Reported the occurrence of failures
    - Reported software used to perform simulation
    - Reported software to perform analysis
3. Justification of data generation
4. Scenarios investigated
5. Statistical methods evaluated
6. Number of simulations performed
7. Justification for number of simulation
8. Criteria to evaluate the performance of statistical methods under different scenarios

For evaluating the statistical methods dealing with MPD the following assessment can be used to assess the performances of the methods: bias, precision or variance, accuracy, type I error, power, and coverage.

Having chosen criteria for our modified checklist, we used them to assess the reporting quality of simulations studies comparing different statistical methods to deal with continuous MPD in RCTs.

Eligibility criteria

We included studies that fulfilled the following criteria:

- Journal articles published in English;
- Articles that discuss statistical methods to how parallel group RCTs might deal with MPD for continuous outcome;
- Articles that compared at least two statistical methods in at least one simulation study;
- Simulation aimed to assess the impact of MPD on estimated treatment effects in RCTs;

We excluded studies that fulfilled any of the following criteria:

- Meeting abstract, letter, commentary, editorials and protocols, books and pamphlets;
- Simulation studies that investigated approaches to dealing with missing data for cluster RCTs, and cross-over RCTs;
- Simulation studies that investigated general performance of methods of imputing missing data but did not focus on its impact on treatment effect in RCTs;
- Methods of handling MPD in health economy studies;
- Duplicate publication

Search strategy

A research librarian with expertise in conducting systematic reviews assisted one of the authors to develop a comprehensive search strategy. We modified and finalized the search strategy using pre-identified eligible articles. We conducted the electronic searches in MEDLINE (from inception to August 2014), Cochrane Library (from inception to August 2014), Web of Science (from inception to January 2015), and Journal Storage (JSTOR) (from inception to January 2015). Appendix 1 presents the detailed search strategy for each database.

Study selection

Two authors independently conducted the title and abstract screening for all the identified citations using a web-based systematic review software (DistillerSR^{TM}). The articles were

included in full-text screening if either reviewer considered the citation might meet the eligibility criteria. The same authors, independently and in duplicate, applied the eligibility

criteria to the full-text of all potentially eligible articles. The reviewers resolved disagreement during full-text screening by discussion or, if unsuccessful, through review by a third author (a statistician).

Data abstraction

We created an Excel spreadsheet to record general characteristics and information related to reporting quality from all eligible simulation studies. General characteristics included last name of the first author, year of publication, missing mechanism investigated, criteria used to evaluate the performance of methods, number of trials simulated, proportion of missing data in the simulated datasets, and number of simulations conducted. We also recorded the clinical area, type of outcome investigated, and, if applicable, number of participants in the trial that motivated the simulation. Reviewers made an assessment of whether each item in the checklist was reported or not reported.

Teams of two authors including one methodologist and one statistician worked in pairs to abstract data independently and in duplicate. Reviewers resolved disagreements by discussion or if necessary through arbitration by a third author. All authors participated in a calibration process for both screening and data abstraction using a pilot form with detailed instructions. The form was modified when the calibration process revealed lack of clarity. For one study, we reviewed supplementary documents published by the

authors to abstract information on reporting quality of simulation when authors referred to it.

Data analysis

We assessed the level of agreement on eligibility between authors for full text screening stage using unweighted kappa[7]. For all analyses, we summarized the categorical variables with numbers and percentages.

Results

The search strategy generated a total of 16,446 citations; 507 were retrieved for full text screening and 60 were deemed eligible (Figure 1). The agreement between authors at the full text screening stage was high (κ= 0.73).

General characteristics of included method studies

Table 1 provides a summary of study characteristics and Table 2 provides a study-by-study detailed description.

Of the 60 studies, 52 (92%) specified the clinical area of the trials that motivated the simulation 18 (30%) investigated infectious diseases and 12 (20%) investigated psychiatry. Approximately half (48.3%) chose a surrogate as the outcomes of interest for their simulations. The total sample size of simulated trials varied from 28[8] to 2000[9,10] with the most common choice being 101-200 (46.7%). Twenty-six (43.3%) studies did not specify the proportion of missing data for the simulated trial. For the studies that did

specify the proportion missing the most common (38.3%) proportion of data missing was 21-30%.

All 60 studies specified at least one criterion to evaluate the performance of statistical methods; 49 (81.7%) assessed bias, 28 (46.7%) assessed coverage, and 26 (43.3%) assessed precision or variance. Over half the studies investigated methods' performance under not missing at random (NMAR) (32, 53.3%) and/or missing at random (MAR) missing mechanisms (32, 53.3%). Three quarters (45, 75%) of the simulations were motivated based on real trials or were described by the authors as typical of real trials.

Reporting quality of simulation studies discussing methods to handle MPD

Table 3 presents a summary of the reporting quality of the eligible studies. Most (54 out of 60: 83.3% with 95% confidence interval [73.9%, 92.7%]) clearly specified the aim of the simulation.

Regarding the simulation procedures, many critical items were not explicitly reported. Most (45, 75% [64.0%, 86.0%]) failed to report whether they created independent simulated datasets for different scenarios. All but one study (59, 98.3% [95%, 101.6%]), failed to report the use of a random number generator (the one that did reported stated they used the random number generator 'normal (0)' in SAS)[11]. No study mentioned their choice of starting seeds, nor whether failure occurred when estimating the outcome of parameter of interest. Over half of the studies reported neither the software package used to perform simulations (35, 58.3% [45.8%, 70.8%]) nor to conduct analysis (31, 51.7% [39.1%, 64.3%]). In those that did provide the information, SAS (13 studies, 21.8%

[11.4%, 32.3%]) was the most frequently applied software both for simulation and for analysis.

A minority (15, 25% [14.0%, 36.0%]) failed to provide justification for data generation. Almost all (57, 95% [89.5%, 100.5%]) reported the number of simulations which varied from 50 to 50,000 replications, with the most common choices being 1000 (23 studies, 38.3% [26.0%, 50.6%]) and equal or less than 250 (12 studies, 20% [9.9%, 30.1%]). Three studies failed to make clear how many simulations were performed[12-14]. Very few (3, 5% [0%, 10.5%]) provided a justification for the number of simulations. Of the 3 that did provide the rationale, two estimated the number of simulation based on an expected standard error around 95% confidence interval of the coverage rate[9,15] and one provided the justification based on the distribution of unstructured covariance matrix[5]. All included studies provided criteria to evaluate the performance of statistical methods.

Discussion

Summary of findings

The reporting quality of published simulation studies addressing methods of handling MPD for individual RCTs suffer from important limitations, especially regarding the simulation procedures. The most egregious omissions included failures during simulation, proportion of missing data, and software to perform simulation or analysis (Table 3). Less serious but frequent omissions included justification of number of simulations, included failure to report the random number generator used, the starting seeds (Table 3).

Strengths and limitations

Our study has several strengths. To our best knowledge, we conducted the first study that systematically assessed the completeness of reporting of these simulation studies. We conducted a systematic and comprehensive search across general medical databases as well as databases that capture statistically oriented articles. The screening and data abstraction processes were performed independently and in duplicate with each pair including one statistician and one methodologist. Chance-corrected agreement in judging eligibility was high. Finally, we used a modification of an established checklist developed specifically for simulation studies to evaluate reporting quality.

One limitation of our study is that we focused exclusively on reporting and did not try to assess the merits of the design and conduct of simulation studies. Therefore, our systematic survey does not explicitly provide information addressing the methodological quality of included simulation studies. Another limitation is that the results might not be generalizable to simulation studies dealing with dichotomous or time to event outcomes, on MPD in non-parallel group RCTs, or on simulation studies addressing issues other than MPD. Lastly, we could have searched EMBASE database to capture potential eligible studies.

Implications

Statisticians and methodologists need to enhance the clarity, completeness and transparency of simulation studies evaluating methods for dealing with MPD for individual RCTs by following standards for comprehensive reporting. Providing explicit

descriptions assists the understanding of readers and makes it more likely that results will be reproducible – or, if they are not, to allow explanations of discrepancies. Transparent reporting reveals drawbacks of research that facilitates the critical appraisal of simulation studies and may play a role in improving the design and conduct of future studies.

Our results suggest that evaluations of other simulation studies may also reveal serious limitations in reporting quality. If this proves a frequent problem, editors of medical and statistical journals may consider endorsing a checklist for the reporting of simulation studies. If such a checklist were adopted and adhered to, poorly reported simulation studies would not pass through the peer review process without correction of the omissions.

Authors' contributions

Conception and design: YZ, TV, LT, GHG

Design of search strategy: EAA, YZ, RC

Paper selection: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, GPM, LECL, FABA, LLCL, JJYN, YF, LW, LK

Data abstraction: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, UA, TI.

Data synthesis: TZ, AA, YZ, YZ

Interpretation of results: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, GPM, LECL, FABA, LCL, JJYN, YF, LW, LK, EAA, GHG

Manuscript drafting: YZ, GHG

Manuscript review and approval: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, GPM, LECL, UA, TI, FABA, LCL, JJYN, YF, LW, LK, DM, EAA, LT, GHG

Funding: This study has no funding support.

Competing interests: All authors declare no conflict of interests.

Figure legends:

Figure 1: PRISMA flow diagram

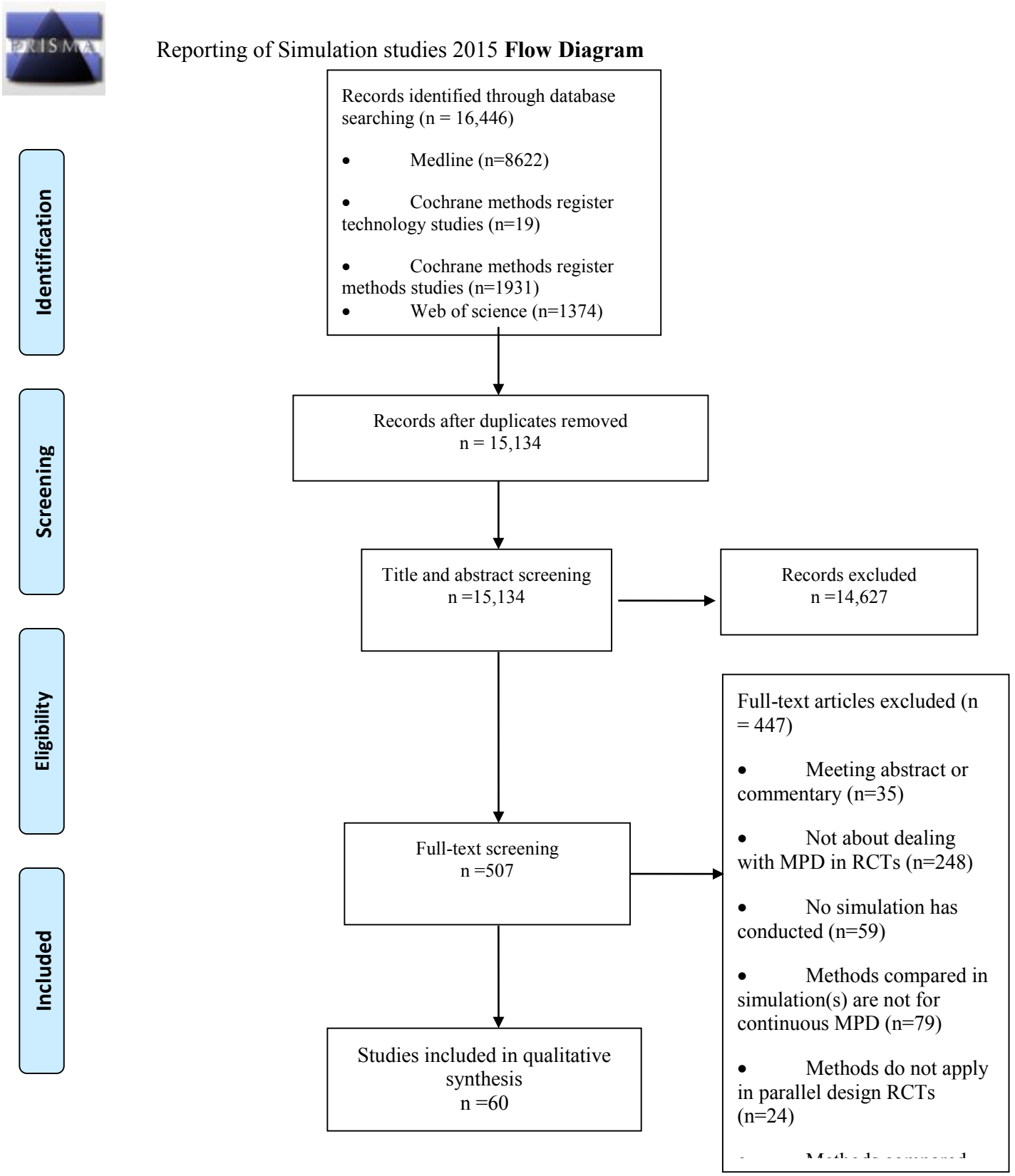

Table 1. Summary of general characteristics of 60 included studies

Clinical Area*	**n**	**%**
Non-Medical	1	1.7
Dermatology	0	0.0
Cardiology	3	5.0
Endocrinology	3	5.0
Gastro Intestinal	1	1.7
Hematology	0	0.0
Intensive Care	0	0.0
Infectious Diseases	18	30.0
Neurology	0	0.0
Oncology	0	0.0
Psychiatry	12	20.0
Renal	1	1.7
Respiratory	3	5.0
Rheumatology	3	5.0
Other	9	15.0
Type of Primary Outcome*	**n**	**%**
Unclear	16	26.7
Length of stay (in hospital, ICU)	1	1.7
Symptoms	5	8.3
Quality of life	3	5.0
Functional status	2	3.3
Disease severity	5	8.3
Length of drug use	3	5.0
Surrogate outcome	29	48.3
Number of different trials	**n**	**%**
1	56	93.3
2	4	6.7
Total Sample size#	**n**	**%**
0-50	1	1.7
50-100	16	26.7
101-200	28	46.7
201-300	12	20.0
301-400	3	5.0
401-500	9	15.0
500-1000	9	15.0
2000	2	3.3

Proportion of missing data*	**n**	**%**
0-4%	1	1.7
5-10%	13	21.7
11%-15%	7	11.7
16%-20%	5	8.3
21%-30%	23	38.3
31%-40%	9	15.0
41%-50%	12	20.0
51%-60%	1	1.7
61%-70%	2	3.3
71%-80%	1	1.7
81%-90%	1	1.7
Unclear	26	43.3
Criteria to assess performance of methods*	**n**	**%**
Bias	49	81.7
Precision/variance	26	43.3
Accuracy	18	30.0
Type I error	11	18.3
Power	13	21.7
Coverage	28	46.7
Missing mechanisms investigated*	**n**	**%**
NMAR	32	53.3
MAR	32	53.3
MCAR	15	25.0
Ignorable missing (MCAR or MAR)	8	13.3
Combined missing (NMAR and MAR)	2	3.3
Software used to perform simulations	**n**	**%**
SAS	13	21.8
R	4	6.7
Splus	3	5.0
C++	2	3.3
STATA	1	1.7
MATLAB	1	1.7
Fortron	1	1.7
Not stated	35	58.3
Software used to perform analysis	**n**	**%**
SAS	19	31.7
Combination	4	6.7
R	2	3.3
C++	2	3.3

Splus	1	1.7
MATLAB	1	1.7
Not stated	31	51.7
Number of simulations	**n**	**%**
Up to 250	12	20.0
400-600	9	15.0
1000	23	38.3
2000	1	1.7
3000	2	3.3
5000	3	5.0
10, 000	5	8.3
25, 000	1	1.7
50, 000	1	1.7
Not stated	3	5.0
Justification for data generation	**n**	**%**
Based on a real data set	32	53.3
Typical of real data	13	21.6
Not stated	15	25.0

NMAR: Not missing at random;
MAR: Missing at random;
MCAR: Missing completely at random.
* The total % of clinical areas, criteria to assess the performance of methods, and missing mechanisms may exceed 100% since there are included studies that simulated more than one trials in different clinical areas, investigating more than one missing mechanism and using more than one criteria to assess the performance of methods.
# The percentage of total sample size may exceed 100% since there are included studies that simulated scenarios with multiple sample sizes.
@ The proportion of missing data may exceed 100% since there are included studies that simulated scenarios with multiple proportion of missing data.

Table2: Characteristics of included studies

Last name of the first author & Year of the publication	Missing mechanism investigated	Criteria used to evaluate the performance of methods: a. A: Bias b. B: Precision c. C: Accuracy d. D: Coverage e. E: Type I error f. F: Power	Number of trials simulated	Clinical application of the simulation	Type of outcome investigated	Number of participants in simulated datasets	Proportion of missing data	Number of simulations
Desouza 2009	MAR	A; F	1	Other	Unclear	100	10% 15% 30%	1000
Horvitz-Lennon 2005	MAR/NMAR	A; C; D	1	Psychiatric	Unclear	500	Not specified	10000
Li 2004	MCAR NMAR	A; C; D	1	Renal	Surrogate outcome	240	Not specified	1000
Lin 2003	MAR NMAR	A; B; D	1	Infectious Diseases	Functional status	200	50% 80% 90%	1000
Peng 2004	MAR NMAR	A; C; D	1	Not specified	Quality of life	200	Not specified	250
Demirtas 2005a	NMAR Ignorable missing	A; B; C; D	1	Infectious Diseases	Surrogate outcome	66	28.78%	1000
Hallgren 2013	MCAR MAR NMAR	A; B; C; E; F	1	Other (Behavioral+ Medication)	Length of drug use	1000	5% 10% 25% 30%	200
Huang 2013	MAR	E; F	1	Rheumatology	Surrogate outcome	100/200/300 /372	50%	1000
Wiens 2013	MAR NMAR	A; E; F	1	Not specified	Unclear	130	10% 25%	10000
Liu 2013	MCAR MAR	A; E; D; F	1	Not specified	Unclear	120/200/300	Not specified	1000
Lepri 1998	NMAR	A	1	Infectious Diseases	Surrogate outcome	100	44%-75%	100
Xue 2011	MAR	B; D	1	Not specified	Unclear	60/100/150/ 250	Not specified	5000
Abrahantes 2011	MAR NMAR	A; C	1	Not specified	Unclear	300	Not specified	100
Wang 1995	MCAR MAR NMAR	A; B; D; F	1	Respiratory	Surrogate outcome	200	50%	600
Hu 2010	MAR	A; B; C	1	Infectious Diseases	Surrogate outcome	500	Not specified	1000
Hedden 2009	MCAR MAR	E; F	1	Other	Length of drug use	100	10% 40%	2000
Liang 2007	MCAR	D	1	Infectious Diseases	Surrogate outcome	200	Not specified	25000
Barnes 2008	NMAR	A; D	1	Psychiatric	Symptoms	60/480	29%	3000
Baron 2008	NMAR	A; E; F	1	Rheumatology	Disease severity	300	8% 15%	1000
Mallinckrodt 2002	NMAR	A; D; E	1	Psychiatric	Symptoms	100	Not specified	3000
Wei 2001	MCAR	A; B; D	1	Respiratory	Length of drug use	300	11%	1000

Roderick 2009	MAR NMAR	A; B; D	1	Infectious Diseases	Surrogate outcome	200	Not specified	1000
Guo 2004	Ignorable missing	A	1	Psychiatric	Symptoms	157	Not specified	500
Carpenter 2007	MAR	A; B	1	Not specified	Unclear	200	50%	1000
Yuan 2007	NMAR	A; B; D	1	Psychiatric	Symptoms	367	Not specified	10000
Tsonaka 2009	NMAR	A; B; C	1	Rheumatology	Symptoms	200	30%	200
Cook 1997	NMAR	A	1	Cardiology	Surrogate outcome	675	NMAR18%+ MAR7%	0
Gueorguieva 2012	I Ignorable missing	A	1	Psychiatric	Quality of life	500	Not specified	200
Hogan 1998	NMAR	F	1	Infectious Diseases	Functional status	424	18% 25%	500
Demirtas 2005b	NMAR Ignorable missing	A; B; D	2	Not specified	Unclear	999	Not specified	1000
Demirtas 2007	NMAR Ignorable missing	A; B; C; D	1	Gastro Intestinal	Surrogate outcome	65	Not specified	1000
Demirtas 2003	NMAR Ignorable missing	B; D	2	Psychiatric and the other one unclear	Unclear and Disease severity	100/413	29%	1000
Longford 2006	NMAR	A; C	1	Psychiatric	Disease severity	28	Not specified	1000
Michael 2012	MAR	Proposed posterior predictive loss model selection criterion	1	Endocrinology	Surrogate outcome	50/100/2000	Not specified	200
Gadbury 2003	MCAR MAR NMAR	A; B; D	1	Obesity	Surrogate outcome	100	20% to 50%	1000
Gilbert 2010	MCAR MAR	A; C; D	1	Infectious Diseases	Surrogate outcome	999	50%	500
Hogan 2004	MCAR	A	1	Infectious Diseases	Surrogate outcome	50	Not specified	0
Groenwold 2011	MAR NMAR	A; B; D	1	Not specified	Unclear	250	30%	5000
James 1995	MAR	A; D	1	Infectious Diseases	Surrogate outcome	500	Not specified	200
Hebert 2011	MAR NMAR	A; B; D	1	Cardiology	Surrogate outcome	416	25%	1000
Lin 2004	MAR	A; B; C	1	Other	Length of stay (in hospital, ICU)	250	Not specified	200
Yi 2009	MAR	A; B; D	1	Infectious Diseases	Surrogate outcome	200/1000	Not specified	500
Stuart 2002	MAR	A	1	Infectious Diseases	Surrogate outcome	200	Not specified	200
Huson 2007	MCAR MAR NMAR	A; B	1	Infectious Diseases	Surrogate outcome	200/500	5% 10% 30%	1000
Wei 1999	MAR	A; D	1	Infectious Diseases	Surrogate outcome	150	12% 26% 29% 33% 37%	500
James 1998	MCAR NMAR	A; D	1	Infectious Diseases	Surrogate outcome	200/2000	Not specified	1000
Lane	MCAR	A; F	1	Psychiatric	Unclear	182/232/244	25%	1000

2008	MAR NMAR						29% 33% 49%	
Li 2011	MCAR	A; B; D	1	Other (vaccine)	Surrogate outcome	200	Not specified	10000
Liu 2006	Ignorable missing	A; C; D	2	Endocrinology	Surrogate outcome	999	22% 24% 41% 42% 48%	0
Mehrotra 2012	MAR	A; E; F	1	Other	Surrogate outcome	60/160	25%-35%	5000
Revicki 2001	NMAR	A	1	Other (physical health status)	Quality of life	200/400/1000	5% 10% 15% 20% 25%	50
Scharfstein 2003	NMAR	B; C	1	Psychiatric	Surrogate outcome	613	34%	500
Siddiqui 2011	Ignorable missing	B; E; F	1	Psychiatric	Unclear	200	35%	1000
Tang 2005	MAR	A; B; D	1	Psychiatric	Disease severity	501	9% 16% 28%	1000
Touloumi 1999	MAR	A; B; C	1	Infectious Diseases	Surrogate outcome	200	Not specified	100
Unnebrink 2001	Combined missing	E; F	1	Endocrinology	Disease severity	266	1% 20%	50000
Wu 2001	MAR	A;C	2	Respiratory and Cardiology	Surrogate outcome	200	11% 10%	400
Xu 2012	MCAR	C; E	1	Infectious Diseases	Surrogate outcome	120	15%-35%	10000
Yuan 2010	NMAR	A; B	1	Infectious Diseases	Surrogate outcome	200	50%	1000
Tseng 2012	MAR NMAR Combined missing	A; C	1	Other	Unclear	200	10%	500

MAR: Missing at random;
NMAR: Not missing at random;
MCAR: Missing completely at random.

Table 3: Summary of reporting quality of included 60 studies

Criteria	n	% (95% CI)
Aims of the simulation		
Reported	50	83.3 [73.9, 92.7]
Dependence of samples		
Samples Independent	15	25.0 [14.0, 36.6]
Starting seed		
Different seeds used	0	0 [0, 3]
Random number generator		
Reported	1	2.0 [0, 5.54]
Failures occur during simulation		
Reported	0	0 [0, 3]
Software to perform simulations		
Reported	25	41.7 [29.2, 54.2]
Software to perform analysis		
Reported	29	48.3 [35.7, 61.0]
Justification for data generation		
Reported	45	75.0 [64.0, 86.0]
Scenarios and statistical methods evaluated		
Reported	60	100.0 [97, 100]
Number of simulations		
Reported	57	95.0 [89.5, 100.5]
Any justification for number of simulations		
Reported	3	5.0 [0, 10.5]
Criteria to evaluate the performance of statistical methods		
Reported	60	100.0 [97, 100]

Appendix 1: Search Strategy (Medline)

#1 "drop-out*".m_titl.

#2 missing.m_titl.

#3 "withdraw*".m_titl.

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

#5 per protocol.m_titl.

#6 intention-to-treat.m_titl.

#7 intent-to-treat.m_titl.

#8 ITT.m_titl.

#9 (exclusion or exclusions).m_titl.

#10 excluded.m_titl.

#11 or/1-11

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

#13 11 and 12

Search Strategy – Cochrane Library

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

Restricted to: Methods Studies and Technology Assessments

Search Strategy – Web of Science

#1 TI=drop-out*

#2 TI=missing

#3 TI=withdraw*

#4 TI=(los* and follow*)

#5 TI=per protocol

#6 TI=intention-to-treat

#7 TI="intention to treat"

#8 TI=intent-to-treat

#9 TI=ITT

#10 TI=(exclusion or exclusions)

#11 TI=excluded

#12 #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1

#13 TI=attrition

#14 #13 OR #12

#15 TS= clinical trial* OR TS=research design OR TS=comparative stud* OR TS=evaluation stud* OR TS=controlled trial* OR TS=follow-up stud* OR TS=prospective stud* OR TS=random* OR TS=placebo* OR TS=(single blind*) OR TS=(double blind*)

#16 #15 AND #12

#17 #15 AND #14

18 TOPIC: (simulat*)

19 #18 AND #17

Search Strategy – JSTOR

((((dropout OR attrition OR missing OR withdraw OR "per-protocol" OR "intent* to treat" OR ITT or "loss to follow*" or "lost to follow") AND ((randomi?ed) or "clinical trial")) AND simulat*) AND la:(eng OR en))

Reference

1. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials.* 2004;1(4):368-376.
2. Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ.* 2012;344:e2809.
3. McKnight PE MK, Sidani S, Figueredo AJ. *Missing Data: A Gentle Introduction.* New York , NY The Guilford Press; 2007.
4. Stubbendick AL, Ibrahim JG. Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics.* 2003;59(4):1140-1150.
5. Cortiñas Abrahantes J, Sotto C, Molenberghs G, Vromman G, Bierinckx B. A comparison of various software tools for dealing with missing data via imputation. *Journal of Statistical Computation and Simulation.* 2011;81(11):1653-1675.
6. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med.* 2006;25(24):4279-4292.
7. TSe HJ. Cochrane Handbook for Systematic Reviews of Interventions (Version 502). 2009
8. Nicholas T. Longford PT, U. A. M. Nur, and H. Seivewright. Analysis of a long-term study of neurotic disorder, with insights into the process of non-response. *J. R. Statist. Soc. A.* 2006;169:507-523.
9. Guo W, Ratcliffe SJ, Have TTT. A Random Pattern-Mixture Model for Longitudinal Data With Dropouts. *Journal of the American Statistical Association.* 2004;99(468):929-937.
10. Daniels MJ, Chatterjee AS, Wang C. Bayesian model selection for incomplete data using the posterior predictive distribution. *Biometrics.* 2012;68(4):1055-1063.

11. Alessandro Cozzi Lepri GDS, Amanda Mocroft, Caroline A. Sabin, Richard W. Morris and Andrew N. Phillips. A practical approach to adjusting for attrition bias in HIV clinical trials with serial marker responses. *AIDS.* 1998;12:1155–1161.
12. Liu M, Wei L, Zhang J. Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. *Pharmaceutical Statistics.* 2006;5(1):7-18.
13. Hogan JW, Lin X, Herman B. Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics.* 2004;60(4):854-864.
14. Cook NR. An imputation method for non-ignorable missing data in studies of blood presure *Statistics in medicine.* 1997;16:2713-2728.
15. Tang L, Song J, Belin TR, Unutzer J. A comparison of imputation methods in a longitudinal randomized clinical trial. *Stat Med.* 2005;24(14):2111-2128.

CHAPTER 3:

HANDLING PARTICIPANTS WITH MISSING OUTCOME DATA FOR CONTINUOUS OUTCOMES IN RANDOMIZED CONTROLLED TRIALS: A SYSTEMATIC SURVEY OF THE METHODS LITERATURE

Yuqing Zhang[1], Akram Alyass[1], Thuva Vanniyasingam[1], Behnam Sadeghirad[1,3], Iván D. Flórez, [1,4] Sathish Chandra Pichika[1], Sean Alexander Kennedy[5], Ulviya Abdulkarimova[6], Yuan Zhang[1], Tzvia Iljon[6], Gian Paolo Morgano[1], Luis E. Colunga Lozano[7], Fazila Abu Bakar Aloweni[8], Luciane C Lopes[9,10], Juan José Yepes-Nuñez[1,11], Yutong Fei[12], Li Wang[13], Lara A Kahale[14], David Meyre[1,2], Elie A Akl[1,14], Lehana Thabane[1], Gordon Guyatt[15]

[1] Department of Clinical Epidemiology and Biostatistics, McMaster University, Canada
[2] Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada
[3] Regional Knowledge Hub and WHO Collaborating Centre for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[4] Department of Pediatrics, University of Antioquia, Medellín, Colombia
[5] Department of Diagnostic Radiology, University of Toronto, Canada
[6] Department of Mathematics and Statistics, McMaster University, Canada
[7] Hospital Civil de Guadalajara, "Fray Antonio Alcalde", México
[8] Singapore General Hospital, Singapore
[9] Universidade de Sorocaba, São Paulo, Brasil
[10] Universidade Estadual Paulista "Julio de Mesquita Filho"São Paulo, Brazil
[11] University of Antioquia, School of Medicine, Medellín, Colombia
[12] Beijing University of Chinese Medicine, McMaster University, China, Canada
[13] Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Canada
[14] Department of Internal Medicine, American University of Beirut, Lebanon
[15] Department of Medicine and Clinical Epidemiology and Biostatistics, McMaster University, Canada

Abstract

Background: Missing participant data (MPD) are a common potential source of bias in randomized controlled trials (RCTs). Appropriate analytic approaches may reduce the bias and produce more trustworthy results. It is likely that some approaches are superior to others.

Objective: To compare the performance of analytical approaches for dealing with MPD for continuous outcomes in individual RCTs.

Method: We conducted a systematic survey of methodological papers indexed on MEDLINE, Cochrane Library, Web of Science, and Journal Storage (JSTOR) up to January 2015. Eligible studies used simulation to compare at least two statistical methods to deal with continuous MPD in RCTs in terms of bias, precision, coverage, accuracy, power, and type I error and overall ranking. We stratified the findings according to the statistical method used and to the pattern of missingness: ignorable (missing completely at random or missing at random) versus non-ignorable (missing not at random).

Result: We included 60 studies comparing 250 methods; 47/60 addressed ignorable and 32/60 non-ignorable data. For both categories, studies addressed a wide variety of methods with limited overlap. The mixed model approach was the most frequently assessed (in 31/60 studies). For each comparison, we report the frequency with which each approach was best followed by the percentage of times tested in which it was the best. For ignorable missing data, mixed model was most frequently the best on overall ranking (9, 34.6%) and bias (10, 55.6%). Multiple imputation also performed well. For non-ignorable data, mixed model was most frequently the best on overall ranking (7,

46.7%), and bias (8, 57.1%). Mixed model performance varied on other criteria including precision and coverage. Last observation carried forward was very seldom the best performing, and for non-ignorable MPD frequently the worst performing method.

Conclusion: The mixed model approach was superior to other methods for dealing with continuous MPD in RCTs, in terms of overall performance and bias. Last observation carried forward performed the worst.

Background

Missing participant data (MPD), broadly defined as "missing information on the phenomena in which we are interested[1]" —also labeled as loss to follow-up, discontinued prematurely or outcome not assessable[2] —is frequent in randomized controlled trials (RCTs). When intervention and control groups have different reasons for MPD, and those reasons are associated with the outcome of interest, the prognostic balance and the validity of the statistical methods used for analysis that randomization is intended to achieve is threatened.

MPD can adversely influence RCT results in 2 ways. Firstly, it may systematically exaggerate or underestimate – that is, bias - the treatment effect. For instance, if there is more likely to be loss to follow-up with worse outcomes in the intervention group than the control group, the treatment effect will be overestimated. Secondly, MPD can reduce the ability of trials to detect true differences between groups (i.e., reduce the statistical power) when only patients with available outcome data are included in the analysis.

Ensuring minimal loss to follow-up is the best approach to deal with MPD. Often, however, despite institution of strategies to minimize MPD, investigators fail to achieve full follow-up in RCTs. A methodological survey found that in a sample of RCTs published in leading general medical journals, 87% reported MPD in their primary binary outcomes[3]. Moreover, the treatment effects in up to one third of the RCTs lost statistical significance when applying plausible assumptions regarding the distribution of events among the MPD in intervention and control groups[3]. It is therefore crucial for clinicians

and researchers to be aware of the risk of bias associated with MPD and for researchers to appropriate apply statistical methods that minimize bias, and to identify the extent to which MPD constitutes risk of bias - whatever the bias-minimizing strategies that are put in place[4].

Statisticians have noted important differences in the mechanism underlying MPD. A commonly used taxonomy proposed by Little and Rubin[5] classifies MPD as missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). If data are truly MCAR, outcomes are identical in distribution in those with MPD and those with complete data, and estimates based on available data will not be misleading. If data are MAR, the probability of being missing is independent of the outcome given the observed values, but available patient characteristics are associated with the outcomes; thus, patient characteristics can be used to make inferences about outcomes in those with MPD. If data are NMAR, missingness is associated with outcomes, and patient characteristics may also associated with missing outcomes, suggesting the necessity for sensitivity analysis. An alternative taxonomy that uses similar concepts refers to ignorable missingness (MCAR or MAR) and non-ignorable (NMAR) missingness[6].

MPD for continuous outcomes provide special challenges[7]. In the past decades, statisticians have proposed many methods to deal with MPD for continuous outcomes in RCTs[8-13]. Common approaches include data deletion, single imputation methods, multiple imputation (MI) methods[14] and data augmentation approaches[15]. Single imputation includes methods such as hot deck, cold deck, mean imputation, regression techniques,

last observation carried forward (LOCF) and composite methods that apply several of the above methods[14].

Single imputation fails to take into account uncertainty in the imputed data and therefore may result in spuriously narrow confidence intervals (CIs)[16,17]. MI builds on the assumption that data in the trials are MAR[1,18-20]. In contrast to single imputation, MI incorporates multiple imputed datasets with consideration of within and between dataset variability that avoids spuriously narrow confidence intervals.

Data augmentation does not explicitly replace missing values. Instead, it invokes an algorithm that takes into account the observed data, the missing data, the relationships among the observed data, and some underlying statistical assumptions when estimating parameters[15]. Common data augmentation methods include model-based approaches such as mixed effects models, robust regression, and s (GEE). These methods are based on maximum likelihood inference[21,22], pseudo-likelihood or maximum inferences[23], and quasi-likelihood inferences[24].

Simulation studies defined as an approach using computer intensive procedure to assess the performance of statistical methods comparing to known truth[25]. Statisticians have used simulation techniques to investigate the relative performance of different methods of dealing with MPD in continuous outcomes[26-31]. In comparison to applying alternative statistical methods to observed data from clinical trials[32,33], simulation has the advantage of assessing performance in relation to the known truth and thus provides more robust evidence of the relative merits of the methods under consideration[25].

We are not aware of any systematic summary of simulation studies evaluating the available statistical methods to deal with continuous MPD in RCTs. We therefore undertook a systematic survey of simulation studies comparing the performance of analytical methods for dealing with continuous MPD in RCTs.

Methods

Definition

We defined MPD as information that is missing for an outcome of interest for a number of trial participants. For these participants, the trialist would typically have information available for their baseline characteristics, and potentially for outcomes other than the one of interest. As a consequence, this study assumes that trialists have access to individual participant data.

MPD does not refer to missing (e.g. unpublished) studies or unreported outcomes (e.g. outcomes planned to report in the protocol but excluded from the trial report).

Eligibility criteria

We included studies that fulfilled all of the following criteria:

- Journal articles published in English;
- Discussed methods for how parallel designed RCTs dealt with continuous MPD as their primary objective;
- Compared at least two approaches in at least one simulation study;

- The simulation study assessed at least one of the following properties: bias, precision, coverage, accuracy, power, and type I error and overall ranking
- Included simulation aimed to assess the impact of MPD on treatment effect in RCTs;

We excluded studies that fulfilled any of the following criteria:

- Meeting abstract, letter, commentary, editorials, protocols, books and pamphlets;
- Missing data not related to individual participant(s) (e.g. missing studies, selective outcome reporting, missing summary data (e.g. SD), missing study level characteristic (e.g. mean age);
- Simulation studies that investigated approaches of handling missing data for cluster RCTs, or cross-over RCTs;
- Simulation studies that handled MPD in health economy studies;
- Simulation studies that investigated general performance of methods of imputing missing data but did not focus on its impact on treatment effect in RCTs;
- Methodological studies summarizing how RCTs reported, dealt with, or judged risk of bias associated with MPD;
- Duplicate publication.

Search strategy

An experienced medical librarian participated in developing the search strategy. We conducted electronic searches in MEDLINE (from inception to August 2014), Cochrane Library (from inception to August 2014), Web of Science (from inception to January

2015), and Journal Storage (JSTOR) (from inception to January 2015). Appendix 1 presents the detailed search strategy for each database.

Study selection

Teams of two reviewers worked in duplicate and independently to screen titles and abstracts of all citations identified in our search. We obtained the full text of all articles that either reviewer deemed as potentially eligible. The same reviewers screened the full texts in duplicate and independently and resolved disagreement through discussion, and when unsuccessful, with the help of a third author (a statistician). We conducted both stages of screening using a web-based systematic review software (DistillerSR^{TM}; https:// systematic-review.ca). For both screening and data abstraction (see details below), we developed and pilot tested standardized forms with clear instructions, and conducted calibration exercises.

Data abstraction

Teams of two reviewers (each including one methodologist and one statistician) abstracted data independently and in duplicate. Teams resolved disagreement through discussion or, if necessary, with assistance from another statistician.

When authors referred to supplementary materials regarding simulation process and the relative performance of methods, we obtained those materials and abstracted the information accordingly.

We used an excel spreadsheet to abstract information related to:

- The general study characteristics
- The missing mechanism(s) of MPD assumed when comparing methods
- The name and type(s) of methods compared in the simulation
- The sample size, overall proportion of missing data, and the distributions used to simulate dataset(s).
- For simulations motivated from clinical trials, we also collected the clinical area of the trial, primary outcome, number of trials simulated.
- For the relative performance of investigated methods, we recorded the ranking of the methods regarding bias, precision, type I error, power, accuracy and coverage, and the overall ranking provided by the authors along with the rationale for the overall ranking.
- When a study investigated multiple factors [25] such as sample size or proportion of missing data that can influence model performance, we recorded the ranking from all scenarios.

Data analysis

Agreement

We assessed agreement between reviewers on full text eligibility using an unweighted kappa. We interpreted kappa values as slight agreement (0.21 to 0.40), moderate agreement (0.41 to 0.60), substantial agreement (0.61 to 0.80), or almost perfect agreement (greater than 0.80)[34].

Classification of findings:

Firstly, we classified results on the basis of the missing mechanism: missing completely at random (MCAR), missing at random (MAR), ignorable missing (either MCAR or MAR), not missing at random (NMAR, also non-ignorable missing), and combined missing (MAR and NMAR) (Figure1).

Based on classification in the literature[13,15,27,28,35] and consensus among authors we created two classification systems for the methods investigated. The first was a 14-category list of approaches including: classic complete case analysis, modified complete case analysis, classic single imputation, modified single imputation, classic LOCF, modified LOCF, classic MI, modified MI, classic mixed model, modified mixed model, classic GEE, modified GEE, classic robust regression, and modified robust regression Categories labelled "classic" referred to approaches conducting analyses while assuming data are based on ignorable missingness, though some can also be applied to data NMAR. When the 14 categories of methods were conceptually similar and performed similarly, we then combined them into an even broader 7-category list of approaches: complete case analysis, single imputation, LOCF, MI, mixed model, GEE, and robust regression. In summaries of method performance in both the 14-category and 7-category classifications, we excluded studies that compared only variations in methods within a single category.

Synthesis of findings:

We recorded rankings (including best and worst) of the performance of categories of methods for each simulation and for each evaluation criterion (bias, precision, type I error,

power, accuracy and coverage) [25] and, if provided by the authors, an overall ranking. The overall ranking typically considered one key property (such as bias) or several properties (e.g. bias and precision, power and accuracy).

We also provided a data summary that combines MCAR, MAR and ignorable missing as labelled by the authors into a category we called "ignorable missing". We decided to collapse the findings in this way because the performance of the methods was similar across above 3 categories as identified by the authors of simulation studies.

Studies investigated the performance of methods using a different number of simulations, each with its own condition(s) (e.g. one study might conduct simulations in 10 conditions that differed in sample size, effect size and proportion of data missing, and another study conduct a single simulation with a single condition). For each mechanism (MCAR, MAR, ignorable missing, NMAR, and combined missing), whether identified by the authors or in our own classification, we counted each study only once.

When studies conducted multiple simulations addressing the same mechanism, and one approach was the best (or the worst) in all simulations, that approach was counted in the summary tables accordingly. When multiple simulations were conducted, and the best or worst approaches varied in simulations, that study was not included in the summary tables. The latter situation occurred for the best performing method, using the authors' classification, in 4 studies for MAR, 3 studies in NMAR and did not occur in MCAR and ignorable missing; for the worst performing method, the situation occurred in 4 studies for MAR, and did not occur in MCAR, ignorable and NMAR.

Results

General characteristics of included studies

Among 16,446 identified citations we retrieved 507 studies for full text screening; 60 proved eligible (Figure 2). The agreement between reviewers for full text screening was substantial (kappa 0.74).

Studies investigated conditions as follows: MCAR (n=15, 25%), and MAR (n=32, 53.3%), ignorable missing (n=8, 13.3%), NMAR (n=32, 53.3%, including 4 reported as non-ignorable missing), and combined missing (n=2, 3.3%) (Table 1). Including MCAR and MAR as ignorable missing, 47 studies investigated ignorable missing data.

The total number of scenarios investigated within each of the 60 studies varied from 1 to 40 with 4 as the most common number of scenarios found in 13 (21.7%) studies.

All studies used at least one criterion to evaluate method performance. Of the 60 studies, 49 (81.7%) assessed bias; 28 (46.7%) coverage; 26 (43.3%) precision; and 42 (70%) provided an overall ranking (Table 1). Appendix 2 presents a study-by-study detailed description.

Of the 60 studies, 52 (92%) specified clinical areas that motivated their studies (either specific clinical trials, or typical of trials in an area, or a reference to a specific clinical context without specific trials). The most common motivating areas were infectious diseases (n=18, 30%) and psychiatry (n=12, 20%) (Table 1). Almost half the studies addressed surrogate outcomes (n=29, 48.3%). The extent of missing data in simulated

trials varied widely (1%[36] to 90%[37]) with the most frequently investigated being 21-30% (n=23, 38.3%). Most studies assumed data were normally distributed (n=38, 63.3%) in at least one of the simulations conducted.

Methods studied

The studies addressed 250 methods. Many authors did not provide the full name for the acronym of methods[38,39], a clear definition of the proposed method[38,40], or official names for proposed methods (instead referring to it as "proposed method")[37,41]. Thus, our classifications of methods required some judgment. Of the 60 studies, the 14-strategy classification system identified 18 (30%) investigated a classic mixed model, 17 (28.3%) modified mixed model; 18 (30%) classic MI; 12 (20%) modified MI; 19 (31.7%) classic LOCF; 4 (6.7%) modified LOCF; 15 (25%) classic complete case analysis; 6 (10%) modified complete case analysis; 15 (25%) classic single imputation; 2 (3.3%) modified single imputation; 8 (13.3%) classic robust regression; 10 (16.7%) modified robust regression; 4 (6.7%) classic GEE; 7 (11.7%) and modified GEE. Of 60 studies, 12 (20.0%) studies compared 2 methods, 16 (26.7%) compared 3 methods, 15 (25.0%) compared 4 methods, 17 (28.3%) compared more than 4 methods (Table 1).

In the 7-categories classification 31 (51.7%) investigated mixed model; 21 (35%) MI; 20 (33.3%) LOCF, 17 (28.3%) complete case analysis, 15 (25%) single imputation, 14 (23.3%) robust regression, and 7 (11.7%) GEE (Table 2). Of the 60 included studies, 20 (33.3%) investigated different methods from only one category, 18 (30%) from two categories, 15 (25%) from three categories, and 7 (11.7%) from four categories (Table 1).

Among 20 studies that investigated only one category of methods, 6 (30%) studies investigated mixed model; 6 (30%) robust regression, 3 (15%) MI, 3 (15%) GEE, and 2 (10%) complete case analysis (Appendix3). Appendix 4 is the map of categories of methods included in each study. Appendix 5 lists all 250 investigated methods with reference number, first author, broad categories, our classifications, methods name and its descriptions.

Performance of included methods

In the following, we first present the 14-category classification of methods, the best performing category for each of MCAR, MAR, ignorable missing, NMAR and combined missing (all as labeled by the authors). We then present, using the 7-category classification, the performance of each method for each of ignorable and non-ignorable MPD (our classification), first with regard to the best approach, then the worst. These summaries are presented as the number of times a method performed best (or worst) and, the percentage in which it was the best (or worst) out of the total times it was compared.

Best performance using 14-category classification

MCAR: Of the 15 (25%) studies that investigated data MCAR, 12 studies compared different categories of methods. Among these 12 studies, 10 (83.3%) reported overall ranking; all reported bias, and 8 (66.7%) precision. Classic mixed model performed the best in the overall ranking (n=4, 80%) and bias (n=4, 100%) (Table 3.1). Classic complete case analysis performed the best second most frequently on bias (n=3, 75%).

Classic LOCF performed the best most frequently regarding precision (n=3, 75%) (Table 3.1).

MAR: Of 32 studies that investigated data MAR, 31 studies compared different categories of methods. Among these 31 studies, 26 (83.9%) reported best overall ranking, 30 (96.8%) bias, 14 (45.2%) precision, 10 (32.2%) accuracy, 10 (32.2%) type I error, 13 (41.9%) power, and 15 (48.4%) coverage. Classic and modified mixed model (n=5, 55.6%; n=4, 57.1%) respectively, performed best in overall ranking, and classic and modified robust regressions (n=3, 60%; n=3, 42.9%) respectively, performed similarly regarding overall ranking; they also performed similarly with respect to bias (Table 3.2). Regarding precision, classic LOCF performed the best most frequently (n=4, 66.7%), and classic robust regression had the highest percentage best (n=3, 100% (Table 3.2). Regarding power, classic LOCF (n=3, 75%) and classic MI (n=3, 75%) were most frequently, and with the highest percentage, the best (Table 3.2). For coverage, classic complete case analysis (n=4, 50%) was the best most frequently and classic MI (n=3, 75%) had the highest percentage best. (Table 3.2)

Ignorable missing data: Of 8 (13.3%) studies that investigated what authors characterized as ignorable missing data, 7 studies compared different categories of methods. Among these 7 studies, 6 (85.7%) reported consistent best overall ranking, 6 (85.7%) bias, 5 (71.4%) precision, 3 (42.9%) accuracy, 1 (14.3%) type I error, 1 (14.3%) power, 4 (57.1%) coverage. Modified MI was most frequently the best on overall ranking (n=2, 50%) and coverage (n=2, 40%). Classic mixed model was also most frequently the best on overall ranking (n=2, 40%), and most frequently on bias (n=3, 100%) and

precision (n=2, 100%). Classic LOCF performed best the most regarding precision (n=2, 100%) and accuracy (n=2, 100%). (Table 3.3)

NMAR: Of 32 (53.3%) studies that investigated NMAR data, 26 studies compared different categories of methods. Among these 26 studies, 24 (92.3%) reported consistent overall ranking, 26 (100%) bias, 11 (42.3%) precision, 11 (42.3%) accuracy, 3 (11.5%) type I error, 6 (23.1%) power, 12 (46.2%) coverage. Classic and modified mixed model performed the best most frequently regarding overall ranking (n=6, 75%; n=3, 42.9%), and on bias (n=6, 85.7%; n=6, 100%) and accuracy (n=4, 80%; n=2, 40%). Classic MI performed best most frequently and with highest percentage on coverage (n=3, 100%). Classic LOCF performed the best most frequently regarding precision (n=4, 57.1%). (Table 3.4)

MAR and NMAR: Two studies[36,42] investigated the combined missing mechanism (MAR and NMAR), there was no clear optimal methods. One study[36] investigated 4 scenarios; classic complete case analysis were the best in 2/4, classic single imputation 1/4 and classic multiple imputation ¼ on overall. The other study[42] investigated 2 scenarios; classic mixed model and modified robust regression performed the best in each scenario on overall ranking and bias respectively.

We observed similar performances of methods in the classic and modified categories across all MCAR, MAR and ignorable missing mechanisms. Classic and modified approaches also performed similarly in the NMAR mechanism. We therefore grouped mechanisms to create ignorable missing (MCAR, MAR, or ignorable) and non-ignorable

missing (NMAR), and combined classic and modified approaches into single broader approaches. For the 14-category system, if an approach (e.g. mixed model) performed similarly in the classic and modified categories (e.g. classic mixed model and modified mixed model, respectively), we presented them both. After combining categories, for such situations in the 7-category summary, we only counted them once. This explains the smaller number of studies in the 7 versus the 14 category summaries.

Best performance using 7-category classification

All types of ignorable missing data: Of the 47 (78.3%) studies that investigated all types of ignorable missing data, 31 (66.0%) studies compared different categories of methods. Among these 31 studies: 21 (67.7%) reported consistent best overall ranking, 22 (71%) bias, 14 (45.2%) precision, 2 (6.4%) accuracy, 4 (12.9%) type I error, 5 (16.1%) power, 14 (45.2%) coverage. Mixed model and MI performed similarly with the highest number best and most frequently the best on overall ranking and bias respectively (n=9, 34.6%; n=8, 38.1%) (n=10, 55.6%; n=7, 43.8%) (Table 4.1). MI performed the best most frequently on coverage (n=6, 50%). LOCF preformed the best most frequently for precision (n=7, 77.8%), and almost always was least frequently the best for the remaining criteria. (Table 4.1)

NMAR: Of 32 (53.3%) studies investigating NMAR data, 23 (71.9%) studies compared different categories of methods. Among these 23 studies: 18 (78.3%) reported best overall ranking, 19 (82.6%) bias, 8 (34.8%) precision, 7 (30.4%) accuracy, 3 (13.0%) type I error, 6 (26.1%) power, 8 (34.8%) coverage. Mixed model performed the best most

frequently on overall ranking (n=7, 46.7%), bias (n=8, 57.1%), accuracy (n=4, 66.7%), and power (n=2, 40%). MI performed approximately as well as mixed model on overall ranking (n=6, 42.9%). MI also performed the best most frequently on coverage (n=4, 50%). LOCF preformed most frequently the best for precision (n=4, 57.1%) (Table 4.2).

Worst performance with 7-categories classification

Studies often presented the best performing methods and infrequently specified the worst performing methods. Therefore, the number of studies presented was much fewer.

For all ignorable missing data, there was little to choose regarding the worst performer across methods (Table 5.1). LOCF was, however, the worst for accuracy (n=2, 66.7%) and coverage (n=2, 28.6%)(Table 5.1).

Among all NMAR simulations, LOCF performed worst most frequently on overall ranking (n=4, 26.7%), bias (n=7, 46.7%); on power (n=3, 75%); and on type I error (n=3, 75%). Complete case analysis had the highest percentage worst on bias (n=4, 57.2%) (Table 5.2). Mixed model was not infrequently the worst on general ranking (n=3, 20%), precision (n=3, 37.5%) and accuracy 2 (33.3%). In all of these 3 cases, mixed model was only compared against robust regression, and robust regression consistently performed better than mixed model in these studies[43-45]. (Table 5.2)

Studies comparing alternatives within single categories of methods

Twenty studies focused on a single category of methods. Because it performed the best of the available methods, we focus here on the studies examining mixed models. Among

6 such studies, 4[35,46-48] investigated ignorable missing data, one study[43] non-ignorable missing data, one study[44] both situations. The sample size of simulated trials varies from 50[47] to 500[46] with number of simulations varies from 200[46,48] to 10,000[43].

A study[44] that compared 6 mixed models found that, when data was NMAR, a lognormal selection model outperformed conditional quadratic models, quadratic/linear model, conditional linear model, and pattern mixture model, regarding overall ranking, bias and accuracy. One study[35] compared random parameter mixture models with shared-parameter model and these 2 methods performed the best regarding bias in different simulations. One study[46] found a joint model with separate dropouts outperformed joint models with common dropout and ignoring dropout on bias in all settings. One study[48] found joint multivariate random effect model outperformed random effect model regarding bias and overall ranking. One study[47] found a varying coefficients model outperformed conditional linear model regarding bias and overall ranking.

Discussion

Main findings

We identified 60 simulation studies that compared 250 methods of dealing with MPD for continuous outcomes in RCTs. Studies addressing both ignorable (MCAR and MAR, 47/60 studies) and non-ignorable (NMAR, 32/60 studies) mechanisms evaluated a wide variety of statistical methods, with limited overlap. The most frequently addressed approach, mixed model, was assessed in 31 studies. Across studies addressing ignorable missing data, mixed model was most frequently the best performing approach on overall

ranking (9), though it was best among all the instances in which it was tested in a third of cases. With respect to bias, mixed model was frequently the highest ranking (10), and also ranked first in a large percentage of instances in which it was best (55.6%), though its performance on other properties (precision, accuracy, Type I error, power, and coverage) was much weaker (Table 4.1). MI also performed well for ignorable MPD (Table 4.1).

Across studies addressing non-ignorable (NMAR) data, mixed model was most frequently ranked best for overall ranking (n=7, 46.7%), bias (n=8, 57.1%), and accuracy (n=4, 66.7%), though seldom on other properties. MI performed similarly well and was second most frequently the best on overall ranking (n=6, 42.9%). Aside from precision, LOCF seldom performed best on any criterion (Table 4.2) and performed the worst most frequently for overall ranking (n=4, 26.7%), bias (n=7, 46.7%), type I error (n=3, 75%) and power (n=3, 75%) (Table 5.2).

Strengths and limitations

Strengths of our study include a comprehensive search, and independent and duplicate screening and data extraction. Pilot testing of data extraction helped ensure the validity of the data collection process. By choosing to summarize only studies comparing statistical methods in simulations for dealing with MPD, we restricted comparisons to those presented in relation to known truth, an approach more compelling than examination of trial results alone, where the truth cannot be known[25].

We examined all the major characteristics relevant to performance of each method. Our pairing of a statistician with a methodologist for data abstraction data helped ensure the accuracy of the process. We applied a strategy to eliminate the cluster effect that would otherwise have occurred if we counted each condition from studies that conducted simulations for several conditions.

Our study has some limitations. The variation in simulation approaches across studies limits strength of inference from our results. Ideally, all studies would have addressed similar criteria (e.g. bias, precision, coverage) using similar assumptions and parameters (such as extent of MPD). Had this been the case, we could have made cross-study comparisons. Because the assessment criteria, statistical assumptions, and parameters differed across studies, we were restricted to within-study comparisons.

Summarizing the results of 60 studies addressing 250 methods proved challenging. First, we had to place methods in categories, a process that involves judgement. Once categorized, summarizing the relative performance of the methods presented challenges. We counted the number of times each method was ranked best or worst, and the percentage of times it was tested best or worst. This approach ignores intermediate performance, and is highly dependent on the number of methods authors chose to compare in individual studies. For instance, one study compared 2 methods and found LOCF was superior to complete case analysis. In this instance, LOCF gets the same credit for being the best as mixed model might when being compared against a number of other methods. The relatively small number of times methods were tested leaves uncertainty about their relative merits. Lastly, due to the variation of assumptions, sample

size of simulated dataset, number of simulations, as well as the criteria used to assess performance of methods, we only summarized the results of studies that evaluated different approaches within mixed model but not other single category.

Interpretation of Findings

The best and worst ranking were generally consistent: methods deemed least frequently to be best were more frequently the worst. For both ignorable and non-ignorable MPD, mixed model and MI were the superior methods with respect to bias. Although MI performed similarly to mixed model, considering it is more complicated to apply in the analysis (multiple data sets need to be analyzed), mixed model is more efficient and to that extent superior[49]. LOCF had very high precision but was worst on bias, type I error and power, which indicated using LOCF would have the estimated effect precisely deviating far from the truth with low power.

Implications

Implications for trialists

Our results suggest trialists should consider using mixed models to deal with MPD whether they believe MPD is or is not ignorable. If they are concerned about minimizing bias, trialists should seldom if ever use LOCF.

Implications for methodologists and future research

Our proposed study provides a comprehensive evidence synthesis on existing approaches and therefore provides direction the development of future research for MPD on continuous outcome in RCTs. Authors proposing new statistical methods should first categorize the methods they are testing. A standard classification system for this categorization would be helpful; in the interval, authors might use the 14-category classification we have proposed. Authors should also provide the full name of the methods with acronyms if applicable using established terminology regarding the name of the methods, and a clear description of the methods. Development of a consensus regarding criteria that define optimal performance of methods of analysis, and statistical procedures for addressing these criteria (e.g. sample sizes and extent of MPD used in simulations) would be highly desirable.

When statisticians choose a mixed model to deal with continuous MPD for trialists, they should consider the empirical results of a simulation study sharing similar characteristics (same missing mechanism, sample size, distribution of the data etc). The 6 simulation studies[35,43,44,46-48] that assessed the performance of the mixed methods - as well as the studies that compared mixed model with other categories of methods – will provide evidence on which to base the selection.

Acknowledgement: We thank Rachel Couban for her assistance regarding design of search strategy and references uploaded with DistillerSR.

Authors' contributions

Conception and design: YZ, TV, AA, EAA, LT, GHG

Design of search strategy: EAA, YZ, Rachel Couban

Paper selection: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, GPM, LECL, FABA, LCL, JJYN, YF, LW, LK

Data abstraction: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, UA, TL.

Data synthesis and analysis: AA, UA, YZ, IDF, TV

Interpretation of results: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, GPM, LC, FABA, LCL, JJYN, YF, LW, LK, EAA, GHG

Manuscript drafting: YZ, GHG

Manuscript review and approval: YZ, AA, BS, TV, IDF, SCP, SAK, YZ, GPM, LECL, UA, TL, FABA, LCL, JJYN, YF, LW, LK, DM, EAA, LT, GHG

Funding: This study has no dedicated funding support.

Competing interests: None of the authors have finial conflict of interests.

Figure legends:

Figure 1: The classification of missing mechanism

	Taxonomy 1	Taxonomy 2
Missing mechanism	**MCAR**	**Ignorable missing**
	MAR	**Ignorable missing**
	NMAR	**Non-ignorable missing**

MCAR: Missing completely at random.
MAR: Missing at random.
NMAR: Not missing at random.

Figure 2: PRISMA flow diagram

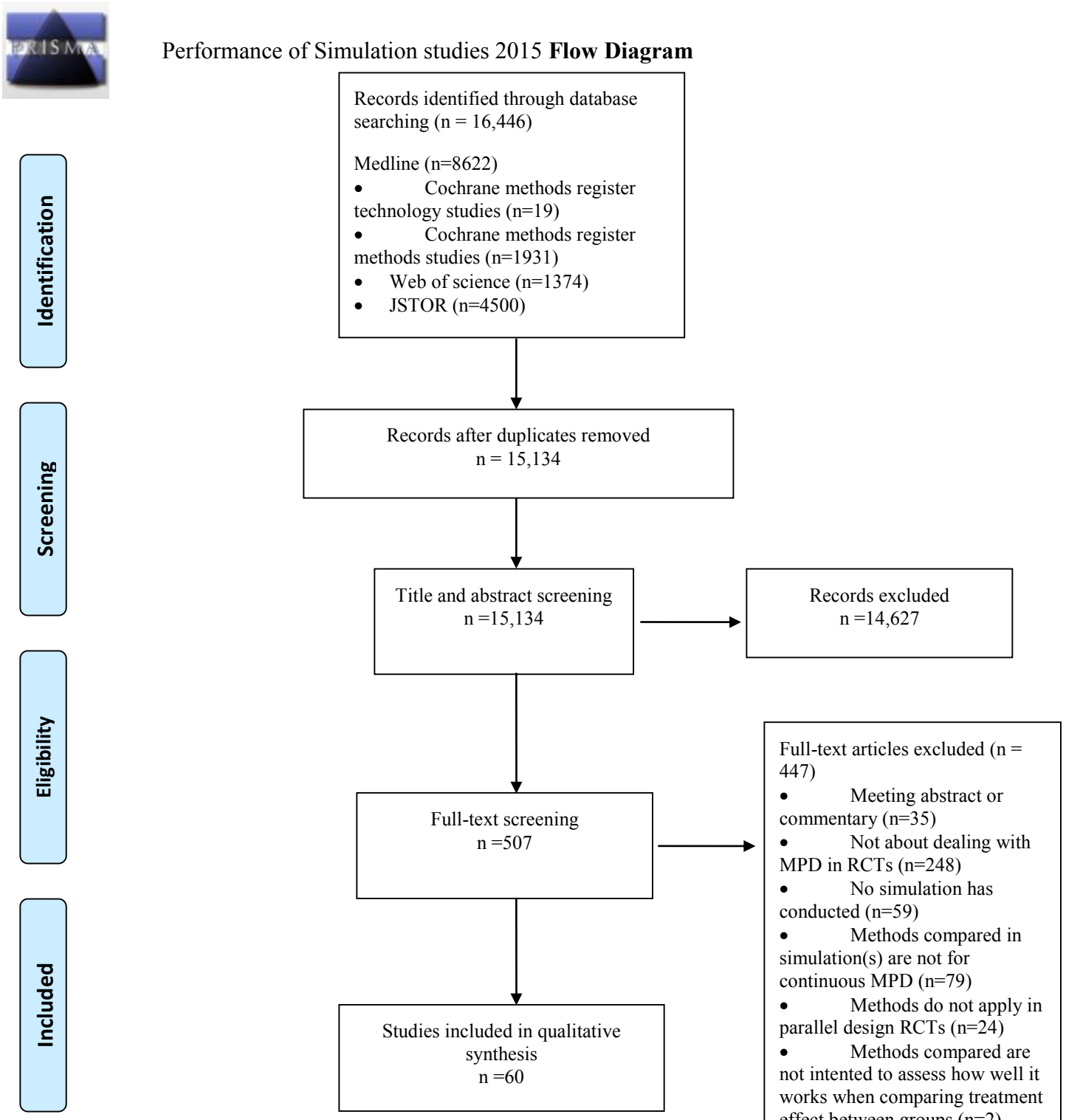

Table 1. Summary of general characteristics of 60 included studies

Clinical Area*	**n (%)**
Non-Medical	1 (1.7)
Cardiology	3 (5.0)
Endocrinology	3 (5.0)
Gastro Intestinal	1 (1.7)
Infectious Diseases	18 (30.0)
Psychiatric	12 (20.0)
Renal	1 (1.7)
Respiratory	3 (5.0)
Rheumatology	3 (5.0)
Other	9 (15.0)
Type of primary outcome	**n (%)**
Unclear	16 (26.7)
Length of stay (in hospital, ICU)	1 (1.7)
Symptoms	5 (8.3)
Quality of life	3 (5.0)
Functional status	2 (3.3)
Disease severity	5 (8.3)
Length of drug use	3 (5.0)
Surrogate outcome	29 (48.3)
Number of different trials simulated	**n (%)**
1	56 (93.3)
2	4 (6.7)
Total sample size#	**n (%)**
0-50	1 (1.7)
50-100	16 (26.7)
101-200	28 (46.7)
201-300	12 (20.0)
301-400	3 (5.0)
401-500	9 (15.0)
500-1000	9 (15.0)
2000	2 (3.3)
Proportion of missing data@	**n (%)**
0-4%	1 (1.7)
5-10%	13 (21.7)
11%-15%	7 (11.7)
16%-20%	5 (8.3)
21%-30%	23 (38.3)
31%-40%	9 (15.0)
41%-50%	12 (20.0)

51%-60%	1 (1.7)
61%-70%	2 (3.3)
71%-80%	1 (1.7)
81%-90%	1 (1.7)
Unclear	26 (43.3)
Number of scenarios investigated	**n (%)**
1	10 (16.7)
2	6 (10.0)
3	9 (15.0)
4	13 (21.7)
5	1 (1.7)
6	6 (6.7)
8	3 (5.0)
9	4 (6.7)
10	1 (1.7)
12	4 (6.7)
15	1 (1.7)
18	1 (1.7)
32	2 (3.3)
40	1 (1.7)
Number of methods investigated	**n (%)**
2	12 (20.0)
3	16 (26.7)
4	15 (25.0)
5	6 (10.0)
6	4 (6.7)
7	2 (3.3)
8	2 (3.3)
11	2 (3.3)
12	1 (1.7)
Number of different categories of methods investigated (based on 7-category classification)	**n (%)**
1	20 (33.3)
2	18 (30.0)
3	15 (25.0)
4	7 (11.7)
Criteria to assess performance of methods*	**n (%)**
Bias	49 (81.7)
Precision	26 (43.3)
Accuracy	18 (30.0)

Type I error	11 (18.3)
Power	13 (21.7)
Coverage	28 (46.7)
Missing mechanisms investigated*	**n (%)**
MCAR	15 (25.0)
MAR	32 (53.3)
Ignorable missing (MCAR or MAR)	8 (13.3)
NMAR	32 (53.3)
Combined missing (NMAR and MAR)	2 (3.3)
Justification for data generation	**n (%)**
Based on a real data set	32 (53.3)
Typical of real data	13 (26.7)
Not stated	15 (25.0)

ICU: Intensive care unit.
MCAR: Missing completely at random.
MAR: Missing at random.
NMAR: Not missing at random.
'methods' refers to the specific method used in each study.
'categories of methods' refers to the 7-category of classification regarding methods.
* The total % of clinical areas may exceed 100% since there are included studies simulated more than one trials in different clinical areas or missing mechanisms.
# The percentage of total sample size may exceed 100% since there are included studies simulated scenarios with multiple sample sizes.
@ The proportion of missing data may exceed 100% since there are included studies simulated scenarios with multiple proportion of missing data.

Table2. Category of methods investigated in 60 included studies*

7-category classification	14-category classification	n (%)
Data deletion	Classic complete case analysis	15 (25)
Data deletion	Modified complete case analysis	6 (10)
Single imputation	Classic single imputation	15 (25)
Single imputation	Modified single imputation	2 (3.3)
Single imputation	Classic LOCF	19 (31.7)
Single imputation	Modified LOCF	4 (6.7)
Multiple imputation	Classic MI	18 (30)
Multiple imputation	Modified MI	12 (20)
Data augmentation	Classic mixed model	18 (30)
Data augmentation	Modified mixed model	17 (28.3)
Data augmentation	Classic GEE	4 (6.7)
Data augmentation	Modified GEE	7 (11.7)
Data augmentation	Classic robust regression	8 (11.7)
Data augmentation	Modified robust regression	10 (16.7)

* The total % of clinical areas may exceed 100% since one study can investigated more than one category of methods
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table 3.1 Best performed methods when reported as MCAR in all scenarios for 14-categories of methods (n=12 studies)*

		n (%)						
Category		**Overall ranking**	**Bias**	**Precision**	**Accuracy**	**Type I error**	**Power**	**Coverage**
Complete case analysis	**Classic**	0	**3 (75)**	1 (33.3)	1 (50)	0	0	0
	Modified	1 (50)	0	0	0	1 (100)	1 (50)	1 (100)
Single imputation	**Classic**	0	1 (25)	0	0	0	0	0
	Modified	0	0	0	0	0	0	0
LOCF	**Classic**	0	1 (20)	**3 (75)**	0	0	1 (33.3)	0
	Modified	1 (50)	1 (50)	1 (50)	0	0	0	0
MI	**Classic**	**2 (40)**	1 (25)	1 (25)	**1 (100)**	**1 (50)**	**1 (50)**	**1 (50)**
	Modified	0	1 (100)	0	0	1 (100)	0	1 (100)
Mixed model	**Classic**	**4 (80)**	**4 (100)**	**1 (25)**	0	**1 (100)**	**1 (50)**	1 (25)
	Modified	**1 (25)**	**1 (25)**	0	**1 (50)**	**1 (50)**	0	**1 (33.3)**
GEE	**Classic**	0	0	0	0	0	0	0
	Modified	0	0	0	0	0	0	1 (100)
Robust regression	**Classic**	1 (100)	0	1 (100)	0	0	0	1 (100)
	Modified	0	0	0	0	0	0	1 (100)

* n is the number of studies that compared between categories of methods.
MCAR: Missing completely at random.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table 3.2 Best performed methods when reported as MAR in all scenarios for 14-categories of methods (n=31 studies)*

		n (%)						
Category		**Overall ranking**	**Bias**	**Precision**	**Accuracy**	**Type I error**	**Power**	**Coverage**
Complete case analysis	**Classic**	1 (10)	2 (25)	0	**2 (66.7)**	1 (50)	0	**4 (50)**
	Modified	2 (40)	0	0	0	1 (100)	1 (50)	1 (25)
Single imputation	**Classic**	1 (12.5)	2 (33.3)	1 (16.7)	0	1 (33.3)	**2 (66.7)**	1 (20)
	Modified	1 (100)	1 (100)	0	0	0	0	1 (100)
LOCF	**Classic**	0	1 (14.3)	**4 (66.7)**	0	1 (33.3)	**3 (75)**	0
	Modified	0	1 (50)	0	0	1 (100)	0	0
MI	**Classic**	2 (18.18)	**3 (37.5)**	2 (28.6)	1 (50)	**2 (66.7)**	**3 (75)**	**3 (75)**
	Modified	3 (50)	2 (33.3)	1 (33.3)	0	**2 (100)**	1 (50)	**2 (66.7)**
Mixed model	**Classic**	**5 (55.56)**	**6 (85.6)**	1 (50)	**2 (100)**	1 (33.3)	1 (25)	0
	Modified	**4 (57.14)**	**5 (83.3)**	1 (33.3)	**3 (100)**	0	**2 (66.7)**	1 (33.3)
GEE	**Classic**	0	0	1 (100)	0	0	0	0
	Modified	1 (20)	1 (20)	0	1 (100)	0	0	0
Robust regression	**Classic**	**3 (60)**	**4 (100)**	**3 (100)**	1 (100)	0	0	0
	Modified	**3 (42.9)**	2(25)	0	0	0	0	**2 (50)**

* n is the number of studies that compared between categories of methods.
MAR: Missing at random.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table3.3 Best performed methods when reported as Ignorable in all scenarios for 14-categories of methods (n=7 studies)*

		n (%)						
Category		**Overall ranking**	**Bias**	**Precision**	**Accuracy**	**Type I error**	**Power**	**Coverage**
LOCF	**Classic**	0	0	**2 (100)**	**2 (100)**	0	0	0
	Modified	0	0	0	0	0	0	0
MI	**Classic**	1 (50)	0	0	0	0	0	0
	Modified	**2 (50)**	1 (25)	1 (25)	0	0	0	2 (40)
Mixed model	**Classic**	**2 (40)**	**3 (100)**	**2 (100)**	0	**1 (100)**	**1 (100)**	0
	Modified	1 (14.7)	**2 (40)**	0	1 (33.3)	0	0	**2 (50)**

***** The remaining 8 classes of methods had 0 count, we therefore omitted those rows.
* n is the number of studies that compared between categories of methods.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table 3.4 Best performed methods when reported as NMAR or non-ignorable missing in all scenarios for 14-categories of methods (n=26 studies)*

		n (%)						
Category		**Overall ranking**	**Bias**	**Precision**	**Accuracy**	**Type I error**	**Power**	**Coverage**
Complete case analysis	**Classic**	0	0	0	0	0	0	1 (16.7)
	Modified	2 (50)	2 (66.7)	0	1 (50)	0	**1 (100)**	**2 (66.7)**
Single imputation	**Classic**	1 (14.3)	2 (28.6)	1 (25)	1 (50)	0	0	0
	Modified	2 (100)	1 (50)	0	0	0	0	0
LOCF	**Classic**	0	1 (7.1)	**4 (57.1)**	0	0	**2 (66.7)**	0
	Modified	0	0	0	0	1 (100)	0	0
MI	**Classic**	**4 (50)**	**4 (50)**	1 (25)	1 (33.3)	1 (50)	1 (50)	**3(100)**
	Modified	2 (25)	0	1 (16.7)	0	0	0	1 (16.7)
Mixed model	**Classic**	**6 (75)**	**6 (85.7)**	1 (33.3)	**4 (80)**	1 (50)	**2 (66.7)**	0
	Modified	**3 (42.9)**	**6 (100)**	0	**2 (40)**	0	0	**3 (50.0)**
GEE	**Classic**	0	0	0	0	0	0	0
	Modified	0	0	0	0	0	0	1 (100)
Robust regression	**Classic**	**2 (66.7)**	**2 (66.7)**	1 (33.3)	**2 (66.7)**	0	0	0
	Modified	**2 (66.7)**	**2 (66.7)**	**2 (66.7)**	0	0	0	1 (100)

* n is the number of studies that compared between categories of methods
NMAR: Not missing at random.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table 4.1 Best performed methods when combine all ignorable missing (reported as MCAR, MAR, or Ignorable missing) in all scenarios for 7-categories of methods (n=31)*

	n (%)						
Category	**Overall ranking**	**Bias**	**Precision**	**Accuracy**	**Type I error**	**Power**	**Coverage**
Complete case analysis	0	1 (11.1)	1 (14.3)	**1 (25)**	**1 (25)**	0	0 (0)
Single imputation	0	1 (16.7)	1 (14.3)	0	0	**1 (33.3)**	2 (33.3)
LOCF	1 (7.7)	2 (18.2)	**7 (77.8)**	0	0	1 (20)	1 (14.3)
MI	**8 (38.1)**	**7 (43.8)**	1 (7.1)	1 (20)	**2 (33.3)**	**2 (28.6)**	**6 (50)**
Mixed model	**9 (34.6)**	**10 (55.6)**	**3 (27.3)**	0	1 (16.7)	1 (11.1)	**4 (33.3)**
GEE	1 (16.7)	0	0	1 (100)	0	0	0
Robust regression	**2 (28.6)**	1 (14.3)	**1 (33.3)**	0	0	0	1 (16.7)

* n is the number of studies that compared between categories of methods.
MCAR: Missing completely at random.
MAR: Missing at random.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table 4.2 Best performed methods when reported as NMAR or non-ignorable missing in all scenarios for 7-categories of methods (n=23)*

	n (%)						
Category	**Overall**	**Bias**	**Precision**	**Accuracy**	**Type I**	**Power**	**Coverage**

	ranking				error		
Complete case analysis	1 (12.5)	3(42.9)	0	0	0	**1 (50)**	1 (20)
Single imputation	3 (42.9)	2(28.6)	1 (25)	1 (50)	0	0	0
LOCF	0	1 (6.7)	**4 (57.1)**	0	1 (25)	**2 (50)**	0
MI	**6 (42.9)**	4 (30.8)	1 (11.1)	1(20)	**1 (50)**	1 (50)	**4 (50)**
Mixed model	**7 (46.7)**	**8 (57.1)**	**1 (12.5)**	**4 (66.7)**	1 (25)	2 (40)	**3 (30)**
GEE	0	1 (50)	0	0	0	0	0
Robust regression	**1 (50)**	**1 (50)**	**1 (50)**	**1 (66.7)**	0	0	0

* n is the number of studies that compared between categories of methods.
NMAR: Not missing at random.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table 5.1 Worst performed methods when combine all ignorable missing (reported as MCAR, MAR, or Ignorable missing) in all scenarios for 7-categories of methods (n=31)*

	n (%)						
Category	**Overall**	**Bias**	**Precision**	**Accuracy**	**Type I**	**Power**	**Coverage**

	ranking				error		
Complete case analysis	0 (0)	1 (11.1)	**2 (28.6)**	2 (40)	0	**1 (25)**	0
Single imputation	1 (10.0)	1 (16.7)	1 (14.3)	1 (50)	0	**1 (33.3)**	1 (16.7)
LOCF	1 (7.7)	1 (9.1)	0	**2 (66.7)**	1 (25)	1 (20)	**2 (28.6)**
MI	0	2 (12.5)	2(14.3)	0	0	0	2 (16.7)
Mixed model	0	0 (0)	2 (18.2)	0	0	2 (22.2)	1 (8.3)
GEE	**1 (20.0)**	**1 (25)**	0	1 (100)	0	0	0
Robust regression	0	1 (14.29)	0	0	0	0	1 (16.7)

* n is the number of studies that compared between categories of methods.
MCAR: Missing completely at random.
MAR: Missing at random.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

Table 5.2 Worst performed methods when reported as NMAR (non-ignorable) in all scenarios for 7-categories of methods (n=23)*

	n (%)

Category	Overall ranking	Bias	Precision	Accuracy	Type I error	Power	Coverage
Complete case analysis	**1 (12.5)**	**4 (57.2)**	**2 (50)**	**1 (50.0)**	1 (100)	**2 (100)**	**2 (40)**
Single imputation	1 (14.3)	2 (28.6)	0	1 (50)	0	0	2 (100)
LOCF	**4 (26.7)**	**7 (46.7)**	0	0	**3 (75)**	**3 (75)**	2 (28.6)
MI	1 (7.1)	5 (38.5)	1 (11.1)	2 (40)	1 (50)	0	**3 (37.5)**
Mixed model	**3 (20.0)**	4 (28.8)	**3 (37.5)**	**2 (33.3)**	0	2 (40)	1 (10.0)
GEE	0	0	0	0	0	0	0
Robust regression	0	1 (50)	0	1 (50.0)	0	0	0

* n is the number of studies that compared between categories of methods.
NMAR: Not missing at random.
LOCF: Last observation caring forward.
MI: Multiple imputation.
GEE: Generalized estimating equations.

References

1. McKnight PE MK, Sidani S, Figueredo AJ. Missing Data: A Gentle Introduction. New York, NY: Guilford; 2007.

2. Akl EA, Kahale LA, Agarwal A, et al. Impact of missing participant data for dichotomous outcomes on pooled effect estimates in systematic reviews: a protocol for a methodological study. *Syst Rev* 2014; **3**: 137.
3. 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.
4. Sampling-based approach to determining outcomes of patients lost to follow-up in antiretroviral therapy scale-up programs in Africa. *JAMA* 2008; **300**(5): 506-7. doi: 10.1001/jama.300.5.506.
5. Little RJA RD. Statistical analysis with missing data. . New York: Wiley; 1987.
6. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med* 2013; **86**(3): 343-58.
7. Thabane L, Mbuagbaw L, Zhang S, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol* 2013; **13**: 92.
8. Demirtas H, Schafer JL. On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Stat Med* 2003; **22**(16): 2553-75.
9. DeSouza CM, Legedza AT, Sankoh AJ. An overview of practical approaches for handling missing data in clinical trials. *J Biopharm Stat* 2009; **19**(6): 1055-73.
10. Hallgren KA, Witkiewitz K. Missing data in alcohol clinical trials: a comparison of methods. *Alcohol Clin Exp Res* 2013; **37**(12): 2152-60.
11. Hebert PL, Taylor LT, Wang JJ, Bergman MA. Methods for using data abstracted from medical charts to impute longitudinal missing data in a clinical trial. *Value Health* 2011; **14**(8): 1085-91.
12. Horvitz-Lennon M, James O'Malley A, Frank RG, Normand S-LT. Improving traditional intention-to-treat analyses: a new approach. *Psychological Medicine* 2005; **35**(7): 961-70.
13. Hu Z, Follmann DA, Qin J. Semiparametric dimension reduction estimation for mean response with missing data. *Biometrika* 2010; **97**(2): 305-19.
14. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol* 2013; **13:92.**(doi): 10.1186/471-2288-13-92.
15. McKnight PE MK SS, Figueredo AJ. Missing Data: A Gentle Introduction. . New York NY The Guilford Press; 2007.
16. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med* 2003; **348**(21): 2136-7.
17. Does analysis using "last observation carried forward" introduce bias in dementia research? *CMAJ* 2008; **179**(8): 751-3. doi: 10.1503/cmaj.080820.
18. Little RJA RD. Tatistical Analysis with Missing Data. 2nd edn ed. New York NY: Wiley; 2002.
19. York RDN. Multiple Imputation for Nonresponse in Surveys. NY: : John Wiley & Sons, Inc; 1987.
20. JL S. Analysis of Incomplete Multivariate Data. New York: Chapman and Hall; 1997.
21. Horton NJ LS. Multiple imputation in practice: com- parison of software packages for regression models with missing variables. *Am Statistician* 2001; **55**: 244–54.

22. G. L. Gadbury CSCaDBA. Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF. *Obes Reb* 2003; **4**(3): 175-84.

23. Liu GF, Hu P, Mehrotra DV. A Two-Step Multiple Imputation for Analysis of Repeated Measures With Left-Censored and Missing Data. *Statistics in Biopharmaceutical Research* 2013; **5**(2): 116-25.
24. Mehrotra DV, Li X, Liu J, Lu K. Analysis of longitudinal clinical trials with missing data using multiple imputation in conjunction with robust regression. *Biometrics* 2012; **68**(4): 1250-9.
25. Burton A AD, Royston P, Holder RL. . The design of simulation studies in medical statistics. *Stat Med* 2006; **25**(24): 4279-92.
26. Stubbendick AL, Ibrahim JG. Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* 2003; **59**(4): 1140-50.
27. Tsonaka R, Verbeke G, Lesaffre E. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics* 2009; **65**(1): 81-7.
28. Xue L, Xue D. Empirical likelihood for semiparametric regression model with missing response data. *Journal of Multivariate Analysis* 2011; **102**(4): 723-40.
29. Yuan Y, Yin G. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* 2010; **66**(1): 105-14.
30. Siddiqui O. MMRM versus MI in dealing with missing data--a comparison based on 25 NDA data sets. *J Biopharm Stat* 2011; **21**(3): 423-36.
31. 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.
32. Bono C, Ried LD, Kimberlin C, Vogel B. Missing data on the Center for Epidemiologic Studies Depression Scale: a comparison of 4 imputation techniques. *Res Social Adm Pharm* 2007; **3**(1): 1-27.
33. Baker SG, Fitzmaurice GM, Freedman LS, Kramer BS. Simple adjustments for randomized trials with nonrandomly missing or censored outcomes arising from informative covariates. *Biostatistics* 2006; **7**(1): 29-40.
34. Higgins JPT TSe. Cochrane Handbook for Systematic Reviews of Interventions (Version 502). Oxford, UK: The Cochrane Collaboration.; 2009
35. Guo W, Ratcliffe SJ, Have TTT. A Random Pattern-Mixture Model for Longitudinal Data With Dropouts. *Journal of the American Statistical Association* 2004; **99**(468): 929-37.
36. Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Stat Med* 2001; **20**(24): 3931-46.
37. LIN DY. Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics* 2003; **4**(3): 385–98.
38. Peng Yahong RJAL, Trivellore E.Raghunathan. An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics* 2004; **60**: 596-607.
39. James M. Robins AR, Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; **90**: 106-21.
40. Xu J, Harrar SW. Accurate mean comparisons for paired samples with missing data: an application to a smoking-cessation trial. *Biom J* 2012; **54**(2): 281-95.
41. Alessandro Cozzi Lepri GDS, Amanda Mocroft, Caroline A. Sabin, Richard W. Morris and Andrew N. Phillips. A practical approach to adjusting for attrition bias in HIV clinical trials with serial marker responses. *AIDS* 1998; **12**: 1155–61.

42. Chi-Hong Tseng RE, Ning Li, Gang Li. Robust inference for longitudinal data analysis with non-ignorable and non-monotonic missing values. *Statistics and its interface*

2012; **5**: 479-90.
43. Yun SC, Lee Y, Kenward MG. Using Hierarchical Likelihood for Missing Data Problems. *Biometrika* 2007; **94**(4): 905-19.
44. Li J, Schluchter MD. Conditional mixed models adjusting for non-ignorable drop-out with administrative censoring in longitudinal studies. *Stat Med* 2004; **23**(22): 3489-503.
45. Yuan Y, Little RJ. Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics* 2009; **65**(2): 478-86.
46. Groenwold RH, Donders AR, Roes KC, Harrell FE, Jr., Moons KG. Dealing with missing outcome data in randomized trials and observational studies. *Am J Epidemiol* 2012; **175**(3): 210-7.
47. Hogan JW, Lin X, Herman B. Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics* 2004; **60**(4): 854-64.
48. Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Impact of missing data due to selective dropouts in cohort studies and clinical trials. *Epidemiology* 2002; **13**(3): 347-55.
49. Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *J Clin Epidemiol* 2013; **66**(9): 1022-8.

Appendix 1: Search Strategy (Medline)

#1 "drop-out*".m_titl.

#2 missing.m_titl.

#3 "withdraw*".m_titl.

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

#5 per protocol.m_titl.

#6 intention-to-treat.m_titl.

#7 intent-to-treat.m_titl.

#8 ITT.m_titl.

#9 (exclusion or exclusions).m_titl.

#10 excluded.m_titl.

#11 or/1-11

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

#13 11 and 12

Search Strategy – Cochrane Library

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

Restricted to: Methods Studies and Technology Assessments

Search Strategy – Web of Science

#1 TI=drop-out*

#2 TI=missing

#3 TI=withdraw*

#4 TI=(los* and follow*)

#5 TI=per protocol

#6 TI=intention-to-treat

#7 TI="intention to treat"

#8 TI=intent-to-treat

#9 TI=ITT

#10 TI=(exclusion or exclusions)

#11 TI=excluded

#12 #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1

#13 TI=attrition

#14 #13 OR #12

#15 TS= clinical trial* OR TS=research design OR TS=comparative stud* OR TS=evaluation stud* OR TS=controlled trial* OR TS=follow-up stud* OR TS=prospective stud* OR TS=random* OR TS=placebo* OR TS=(single blind*) OR TS=(double blind*)

#16 #15 AND #12

#17 #15 AND #14

18 TOPIC: (simulat*)

19 #18 AND #17

Search Strategy – JSTOR

((((dropout OR attrition OR missing OR withdraw OR "per-protocol" OR "intent* to treat" OR ITT or "loss to follow*" or "lost to follow") AND ((randomi?ed) or "clinical trial")) AND simulat*) AND la:(eng OR en)

Appendix2: Characteristics of included studies

Appendix2: Characteristics of included studies

Last name of the first author & Year of the publication	Missing mechanism investigated	Criteria used to evaluate the performance of methods: g. A: Bias h. B: Precision i. C: Accuracy j. D: Coverage k. E: Type I error l. F: Power	Number of trials simulated	Clinical application of the simulation	Type of outcome investigated	Number of participants in simulated datasets	Proportion of missing data	Number of simulations
Desouza 2009	MAR	A; F	1	Other	Unclear	100	10% 15% 30%	1000
Horvitz-Lennon 2005	MAR/NMAR	A; C; D	1	Psychiatric	Unclear	500	Not specified	10000
Li 2004	MCAR NMAR	A; C; D	1	Renal	Surrogate outcome	240	Not specified	1000
Lin 2003	MAR NMAR	A; B; D	1	Infectious Diseases	Functional status	200	50% 80% 90%	1000
Peng 2004	MAR NMAR	A; C; D	1	Not specified	Quality of life	200	Not specified	250
Demirtas 2005a	NMAR Ignorable missing	A; B; C; D	1	Infectious Diseases	Surrogate outcome	66	28.78%	1000
Hallgren 2013	MCAR MAR NMAR	A; B; C; E; F	1	Other (Behavioral+ Medication)	Length of drug use	1000	5% 10% 25% 30%	200
Huang 2013	MAR	E; F	1	Rheumatology	Surrogate outcome	100/200/300 /372	50%	1000
Wiens 2013	MAR NMAR	A; E; F	1	Not specified	Unclear	130	10% 25%	10000
Liu 2013	MCAR MAR	A; E; D; F	1	Not specified	Unclear	120/200/300	Not specified	1000
Lepri 1998	NMAR	A	1	Infectious Diseases	Surrogate outcome	100	44%-75%	100
Xue 2011	MAR	B; D	1	Not specified	Unclear	60/100/150/ 250	Not specified	5000
Abrahantes 2011	MAR NMAR	A; C	1	Not specified	Unclear	300	Not specified	100
Wang 1995	MCAR MAR NMAR	A; B; D; F	1	Respiratory	Surrogate outcome	200	50%	600
Hu 2010	MAR	A; B; C	1	Infectious Diseases	Surrogate outcome	500	Not specified	1000
Hedden 2009	MCAR MAR	E; F	1	Other	Length of drug use	100	10% 40%	2000
Liang 2007	MCAR	D	1	Infectious Diseases	Surrogate outcome	200	Not specified	25000
Barnes 2008	NMAR	A; D	1	Psychiatric	Symptoms	60/480	29%	3000
Baron 2008	NMAR	A; E; F	1	Rheumatology	Disease severity	300	8% 15%	1000
Mallinckrodt 2002	NMAR	A; D; E	1	Psychiatric	Symptoms	100	Not specified	3000
Wei 2001	MCAR	A; B; D	1	Respiratory	Length of drug use	300	11%	1000

Roderick 2009	MAR NMAR	A; B; D	1	Infectious Diseases	Surrogate outcome	200	Not specified	1000
Guo 2004	Ignorable missing	A	1	Psychiatric	Symptoms	157	Not specified	500
Carpenter 2007	MAR	A; B	1	Not specified	Unclear	200	50%	1000
Yuan 2007	NMAR	A; B; D	1	Psychiatric	Symptoms	367	Not specified	10000
Tsonaka 2009	NMAR	A; B; C	1	Rheumatology	Symptoms	200	30%	200
Cook 1997	NMAR	A	1	Cardiology	Surrogate outcome	675	NMAR18%+ MAR7%	0
Gueorguieva 2012	I Ignorable missing	A	1	Psychiatric	Quality of life	500	Not specified	200
Hogan 1998	NMAR	F	1	Infectious Diseases	Functional status	424	18% 25%	500
Demirtas 2005b	NMAR Ignorable missing	A; B; D	2	Not specified	Unclear	999	Not specified	1000
Demirtas 2007	NMAR Ignorable missing	A; B; C; D	1	Gastro Intestinal	Surrogate outcome	65	Not specified	1000
Demirtas 2003	NMAR Ignorable missing	B; D	2	Psychiatric and the other one unclear	Unclear and Disease severity	100/413	29%	1000
Longford 2006	NMAR	A; C	1	Psychiatric	Disease severity	28	Not specified	1000
Michael 2012	MAR	Proposed posterior predictive loss model selection criterion	1	Endocrinology	Surrogate outcome	50/100/2000	Not specified	200
Gadbury 2003	MCAR MAR NMAR	A; B; D	1	Obesity	Surrogate outcome	100	20% to 50%	1000
Gilbert 2010	MCAR MAR	A; C; D	1	Infectious Diseases	Surrogate outcome	999	50%	500
Hogan 2004	MCAR	A	1	Infectious Diseases	Surrogate outcome	50	Not specified	0
Groenwold 2011	MAR NMAR	A; B; D	1	Not specified	Unclear	250	30%	5000
James 1995	MAR	A; D	1	Infectious Diseases	Surrogate outcome	500	Not specified	200
Hebert 2011	MAR NMAR	A; B; D	1	Cardiology	Surrogate outcome	416	25%	1000
Lin 2004	MAR	A; B; C	1	Other	Length of stay (in hospital, ICU)	250	Not specified	200
Yi 2009	MAR	A; B; D	1	Infectious Diseases	Surrogate outcome	200/1000	Not specified	500
Stuart 2002	MAR	A	1	Infectious Diseases	Surrogate outcome	200	Not specified	200
Huson 2007	MCAR MAR NMAR	A; B	1	Infectious Diseases	Surrogate outcome	200/500	5% 10% 30%	1000
Wei 1999	MAR	A; D	1	Infectious Diseases	Surrogate outcome	150	12% 26% 29% 33% 37%	500
James 1998	MCAR NMAR	A; D	1	Infectious Diseases	Surrogate outcome	200/2000	Not specified	1000
Lane	MCAR	A; F	1	Psychiatric	Unclear	182/232/244	25%	1000

2008	MAR NMAR						29% 33% 49%	
Li 2011	MCAR	A; B; D	1	Other (vaccine)	Surrogate outcome	200	Not specified	10000
Liu 2006	Ignorable missing	A; C; D	2	Endocrinology	Surrogate outcome	999	22% 24% 41% 42% 48%	0
Mehrotra 2012	MAR	A; E; F	1	Other	Surrogate outcome	60/160	25%-35%	5000
Revicki 2001	NMAR	A	1	Other (physical health status)	Quality of life	200/400/1000	5% 10% 15% 20% 25%	50
Scharfstein 2003	NMAR	B; C	1	Psychiatric	Surrogate outcome	613	34%	500
Siddiqui 2011	Ignorable missing	B; E; F	1	Psychiatric	Unclear	200	35%	1000
Tang 2005	MAR	A; B; D	1	Psychiatric	Disease severity	501	9% 16% 28%	1000
Touloumi 1999	MAR	A; B; C	1	Infectious Diseases	Surrogate outcome	200	Not specified	100
Unnebrink 2001	Combined missing	E; F	1	Endocrinology	Disease severity	266	1% 20%	50000
Wu 2001	MAR	A;C	2	Respiratory and Cardiology	Surrogate outcome	200	11% 10%	400
Xu 2012	MCAR	C; E	1	Infectious Diseases	Surrogate outcome	120	15%-35%	10000
Yuan 2010	NMAR	A; B	1	Infectious Diseases	Surrogate outcome	200	50%	1000
Tseng 2012	MAR NMAR Combined missing	A; C	1	Other	Unclear	200	10%	500

MAR: Missing at random;
NMAR: Not missing at random;
MCAR: Missing completely at random.

Appendix 3: Number of categories of methods investigated

Appendix 3.1: Studies investigated more than one category of methods

	All	Ignorable (n=31)	NMAR (n=23)
Complete case analysis	15	13	7
Single imputation	15	11	7
LOCF	20	13	15
MI	23	18	13
Mixed model	25	21	15
GEE	4	3	1
Robust regression	8	7	2

* NMAR: not missing at random.

Appendix3.2: Studies investigated one category of method

	All	Ignorable (n=16)	NMAR (n=9)
Complete case analysis	2	1	2
Single imputation		0	0
LOCF		0	0
MI	3	3	1
Mixed model	6	5	2
GEE	3	3	1
Robust regression	6	4	3

* NMAR: not missing at random.

Appendix 4: Categories of methods included in each study

Last name of the first author & Year of the publication	Complete case analysis		Single imputation		Last observation carrying forward		Multiple imputation		Mixed model		Generalize estimate equation		Robust regression	
	C	M	C	M	C	M	C	M	C	M	C	M	C	M
Desouza 2009							✓		✓		✓	✓		
Horvitz-Lennon 2005	✓	✓												
Li 2004										✓				
Lin 2003													✓	✓
Peng 2004	✓	✓							✓	✓				
Demirtas 2005a	✓				✓			✓	✓	✓				
Hallgren 2013			✓		✓		✓			✓				
Huang 2013	✓		✓		✓		✓							
WIENS 2013	✓					✓			✓					
Liu 2013							✓	✓		✓				
lepri 1998			✓	✓	✓									
Xue 2011	✓		✓										✓	
Abrahantes 2011							✓	✓						
Wang 1995	✓	✓							✓	✓				
Hu 2010													✓	✓
Hedden 2009		✓	✓				✓		✓					
Liang 2007			✓		✓								✓	
Barnes 2008			✓				✓		✓					
Baron 2008	✓						✓		✓					
Mallinckrodt 2002					✓	✓			✓	✓				
Wei 2001	✓				✓	✓			✓					
Roderick 2009					✓				✓					
Guo 2004									✓	✓				
Carpenter 2007							✓	✓						
Yuan 2007										✓				
Tsonaka 2009										✓				
Cook 1997			✓		✓			✓						
Gueorguieva 2012									✓	✓				
Hogan 1998		✓												
Demirtas 2005b								✓						
Demirtas 2007					✓			✓		✓				
Demirtas 2003								✓	✓	✓				
Longford 2006			✓		✓		✓							
Michael 2012									✓	✓			✓	
Gadbury 2003	✓				✓		✓		✓					
GILBERT 2010	✓													✓
Hogan 2004									✓					
Groenwold 2011	✓		✓					✓						
James 1995											✓	✓		
Hebert 2011			✓		✓		✓							
Lin 2004											✓	✓		
Yi 2009												✓		
Stuart 2002									✓	✓				
Huson 2007			✓		✓	✓	✓							
Wei 1999													✓	✓
James 1998												✓		
Lane 2008					✓				✓					
Li 2011														
Liu 2006					✓		✓	✓	✓	✓				
Mehrotra 2012								✓	✓		✓	✓		
Revicki 2001			✓	✓	✓									
Scharfstein 2003														✓
Siddiqui 2011							✓		✓					
Tang 2005	✓		✓		✓		✓							

Touloumi 1999									✓	✓				
Unnebrink 2001	✓		✓		✓		✓							
Wu 2001														✓
Xu 2012	✓		✓											
Yuan 2010	✓												✓	✓
Tseng 2012									✓					✓

C: Classic methods; M: Modified methods.

Appendix 5: All included methods

ID	Author	Classic vs. not	Method	Description
17	DeSouza	Classsic mixed model	Mixed effects model for repeated measurements (MMRM)	Method is based on maximum likelihood. Assumptions include i)an unstructured mean response across visits, ii)a variance–covariance structure with an AR(1) correlation between repeated measures, iii) subjects as a random effect, and iv) constant residual variability at
17	DeSouza	Classic GEE	Unweighted generalized estimating equations (GEE) method	A semi-parametric regression-based strategy that requires specification of the structure of means and variances, and a working correlation for the multivariate measurements. Distribution of the
17	DeSouza	Modified GEE	Weighted generalized estimating equations (WGEE) method	A weight is assigned at the participant level, and calculated as the inverse probability for dropping out at the observed time of dropout. The probability of a dropout is based on a logistic regression model for
17	DeSouza	Classic multiple imputation	Multiple-imputation-based GEE (MI-GEE)	Values are drawn from the conditional distribution of the unobserved responses given the observed responses and observed covariates. Imputed datasets are analyzed using standard analysis methods, and
43	Hovirtz-Len	Classic complete case analysis	Complete case	Keep and analyze only individuals with complete data.
43	Hovirtz-Len	Classic complete case analysis	Complete case analysis post removal of non-compliant subjects (ITT analysis using weighted observation analysis)	For trials with non-compliance, the general approach is to first infer who is a latent complier and then estimate the treatment effect for this sub- population. This may be accomplished by ex- pressing the ITT estimate for the entire sample as a function involving (a) the ITT effects of assigned treatment for each compliance state and (b) the
43	Hovirtz-Len	Modified complete case analysis	Stratified method of moments (SMOM)	The new estimator, referred to as a stratified method of moments (SMOM) estimator, has good statistical properties. It closely estimates the true treatment effect, that is, it is unbiased, and it approaches the
57	Li	Modified mixed model	Conditional quadratic models	Slopes and intercepts follow quadratic regression curves on drop-out or censoring time, stratified according to whether or not drop-out
57	Li	Modified mixed model	Quadratic/linear model	No specific definition provided
57	Li	Modified mixed model	Conditional linear model	Slopes and intercepts follow linear regression curves on drop-out or censoring time, stratified according to whether or not drop-out
57	Li	Modified mixed model	Pattern mixture model	When length of follow-up is the same for all subjects who do not drop out, pattern mixture models can be used. A general mixed-model framework, where groups are set by distinct categories according to
57	Li	Modified mixed model	Ignorable model	Mixed model consider non-ignorable drop out
57	Li	Modified mixed model	Lognormal selection model	No specific definition provided
61	Lin	Classic robust regression	Proposed semi-parametric regression:	This model is more flexible than the conventional parametric models for longitudinal data in that the evolution of the response over time,
61	Lin	Modified robust regression	Naïve semi-parametric regression	Naive method of Lin and Ying (2001
85	Peng	Classic complete case analysis	Naïve instrumental variable approach (IVn)	Unclear
85	Peng	Modified complete case analysis	Instrumental variable approach	The estimator applied to the complete cases
85	Peng	Modified complete case analysis	Instrumental variable approach with naïve latent esimates	No specific definition provided; (corresponding extended IV estimators assuming latent ignorable mechanisms)
85	Peng	Modified complete case analysis	Instrumental variable approach with latent estimates	No specific definition provided
85	Peng	Modified mixed model	Maximum likelihood before deletion	Maximum likelihood estimate of the CACE of the data before generating msissing values; CACE is the complier-average causal effect for the subpopulation of individuals who would comply under either
85	Peng	Classic mixed model	MLI	Maximum likelihood of CACE assuming ignorable mechanisms; CACE is the complier-average causal effect for the subpopulation of
85	Peng	Classic mixed model	BYSLI (acronym not defined in paper)	Bayesian estimates of CACE assuming ignorable mechanisms; CACE is the complier-average causal effect for the subpopulation of
85	Peng	Modified mixed model	MLLI (acronym not defined in paper)	Maximum likelihood of CACE assuming latent ignorable mechanisms; CACE is the complier-average causal effect for the subpopulation of individuals who would comply under either treatment
85	Peng	Modified mixed model	BYSLI (acronym not defined in paper)	Bayesian estimates of CACE assuming latent ignorable mechanisms; CACE is the complier-average causal effect for the subpopulation of individuals who would comply under either treatment
85	Peng	Modified mixed model	MLCC (acronym not defined in paper)	Maximum likelihood estimate of the CACE of the data after deleting observations with missing values; CACE is the complier-average causal effect for the subpopulation of individuals who would comply under

85	Peng	Classic mixed model	BYSCC (acronym not defined in paper)	Bayesian estimates of the CACE of the data after deleting observations with missing values; CACE is the complier-average causal effect for the subpopulation of individuals who would comply under either
138	Demirtas	Classic LOCF	LOCF	The last available measurement is carried forward to fill in unobserved
138	Demirtas	Classic complete case analysis	Complete case analysis	Subjects having full set of measurements are considered for the analysis. A linear mixed model is used.
138	Demirtas	Classic maximum likelihood	Maximum likelihood under a linear mixed-effects model	Direct maximum likelihood is used under a linear mixed-effects model.
138	Demirtas	Modified mixed model	OSWALD-IGN using OSWALD software package	Assumes dropouts are ignorabeand dropouts depend on the wave of the measurements and the previous responses. A linear mixed model
138	Demirtas	Modified mixed model	OSWALD-NIGN using OSWALD software package	Assumes dropouts depend on the wave of the measurement and the current response. A linear mixed model is used.
138	Demirtas	Modified multiple imputation	Conventional pattern mixture model with a linear component (PMM-LIN)	A pattern mixture model that uses a linear term for time of last observation and its interactions with drug indicator and weeks of measurement. MCMC routines is used to create multiple imputations.
138	Demirtas	Modified multiple imputation	Conventional pattern mixture model with a dummy indicator component (PMM-DROP)	A pattern-mixture model that uses a dummy indicator of completion status and its interactions with drug indicator and weeks of measurement. MCMC routines is used to create multiple imputations.
138	Demirtas	**Modified multiple imputation**	Hierarchical Bayesian pattern-mixture model (HBPMM)	A hierarchical Bayesian pattern-mixture model that allows decreasing amount of variability among high order time polynomials. MCMC routines is used to create multiple imputations.
201	Hallgren	Classic single imputation	Worse case scenario (WCS)	Participants with missing data were assumed to have relapsed to daily heavy drinking (100% PHD).
201	Hallgren	Classic multiple imputation	Multiple Imputation method	Datasets created with missing values within their plausible ranges. Imputation performed using chained regression equations where imputed values are estimated based on the mean and covariance structures among the missing data variable and the set of auxiliary
201	Hallgren	Modified mixed model	Full information maximum likelihood (FIML)	All available data and the maximum likelihood (using an expectation maximization algorithm) are used to identify the values of the model parameters that maximize the fit of the model to the observed data. Maximum likelihood selects parameter estimates that minimize the
201	Hallgren	Classic LOCF	LOCF	Values from the most recently available time point were used to
208	Huang	Classic complete case analysis	Completers	Complete case analysis
208	Huang	Classic multiple imputation	Mean rank imputation	A novel imputation method that uses rank transformation for missing data imputation. Convert original progression scores into ranks, where all missing values are imputed with a rank. A mean rank is imputed
208	Huang	Classic single imputation	Linear extrapolation	Assumes a linear relationship exists between progression and time to impute values for patients missing 1-year progression scores with twice the amount of their 6-month progression scores.
208	Huang	Classic single imputation	Median quartile bin imputation	Imputes missing values of the outcome with the median of the observed 1-year data from the respective quartile bin defined by the
208	Huang	Classic single imputation	Median imputation	Imputes the missing 1-year progression scores (outcome) with the median of all observed 1-year data.
208	Huang	Classic multiple imputation	Mean rank imputation	(nonparametric approach) 3, 5, 7, and 9 sets imputed per sample
208	Huang	Classic LOCF	LOCF	carries forward the last observed progression score to 1 year
209	Wiens	Classic complete case analysis	All available analysis	Same as above definition
209	Wiens	Modified LOCF	Last observation carried forward(LOCF-DELTA)	Same as above definition
209	Wiens	Classic mixed model	Mixed effect model	Mixed-effect model repeated measured (MMRM) with an unstructured covariance matrix
247	Liu	Modified multiple imputation	Conventional single imputation	This approach is to handle left censored data. It first replaces left censored data with half of the detection limit and then analyzes the entire data with constraint longitudinal data analysis (cLDA) models.
247	Liu	Modified multiple imputation	Two-step MI followed by adaptive method	A two-step MI procedure is investigated for left-censored observations and missing data. Imputation is performed by fitting a normal distribution. Twenty sets are imputed per sample with two potential methods of analysis (adaptive method). For each imputed dataset, the scaled residuals from the cLDA model, are tested for normality using adjusted p-values. If p-values are less than 0.001, then normality is rejected and the last time point values are analyzed with a
247	Liu	Modified mixed model	Regression model using NLMIXED in SAS	The NLmixed model is a random effects model that is fit on data from the baseline and at the last time points. The marginal distributions are assumed to be normal where the baseline means of two treatment

247	Liu	Classic multiple imputation	Two-step MI followed by constraint longitudinal data analysis (cLDA)	Left-censored data and missing data are first imputed with a two-step MI, 20 sets per dataset. For each completed dataset, a cLDA model is then created to fit the baseline and last time point values. Treatment, time as a categorical variable, and treatment by time interaction; and an unstructured covariance can all be investigated.
346	Lepri	**Classic LOCF**	Last observation carried forward	No specific definition provided
346	Lepri	**Modified single imputation**	Other Single Imputation	Adoption of future value(s) from a patient whose change in a specific variable is closest to the value at the tie of drop-out
346	Lepri	**Classic single imputation**	Graphical approach	Plot the mean changes of the outcome for patients receiving treatment. Average values are used for imputation.
385	Xue	Classic single imputation	Normal approximation to estimate β_1	An estimator (β_1) is assumed to be asymptotically normal
385	Xue	Classic complete case analysis	Normal approximation to estimate β_c	An estimator (βc) is assumed to be asymptotically normal
385	Xue	Classic robust regression	Complete case empirical likelihood (CEL)	Using complete case data; used semiparametric regression imputation to construct empirical likelihoods. Empirical likelihood is a nonparametric or semi parametric method used for deriving statistical
385	Xue	Classic single imputation	Imputed empirical likelihood (IEL)	Using imputed data; used semiparametric regression imputation to construct empirical likelihoods. Empirical likelihood is a nonparametric or semi parametric method used for deriving statistical estimators.
385	Xue	Classic complete case analysis	Adjusted empirical likelihood (AEL)	Unclear description; used semiparametric regression imputation to construct empirical likelihoods. Empirical likelihood is a nonparametric or semi parametric method used for deriving statistical estimators.
385	Xue	Classic single imputation	Weight-corrected empirical likelihood (WCEL)	Unclear description; used semiparametric regression imputation to construct empirical likelihoods. Empirical likelihood is a nonparametric or semi parametric method used for deriving statistical estimators.
405	Abrahantes	Classic multiple imputation	Multiple imputation using SEM algorithm	Stochastic expectation maximization (SEM) algorithm is an iterative algorithm used for calculating maximum likelihood estimates. Missing data are imputed with plausible values, using observed values,
405	Abrahantes	Classic multiple imputation	Multiple imputation using Random forest approach	Many identically distributed trees are generated using bootstrapping to create a proximity matrix. Proximities are matched with complete data, normalized, and proximity measures are then used to impute
405	Abrahantes	Classic multiple imputation	Multiple imputation using PROC MI in Software SAS	Three methods (MCMC approach, regression method, and propensity score method) can be used to incorporate appropriate variability across the M imputations. The method selected depends on the
405	Abrahantes	Modified multiple imputation	Multiple imputation using Hmisc	Hmisc is a package in statistical software R. Multiple imputation is performed using additive regression, bootstrapping, and predictive
405	Abrahantes	Modified multiple imputation	Multiple imputation using Amelia II packages from R	An alogorithm based on bootstrapping. A sample is bootstrapped, statistics are estimated using priors, and missing values are imputed.
418	Wang	Classic complete case analysis	All available Data	No additional information provided
418	Wang	Modified complete case analysis	Single imputation by Regression	No additional information provided
418	Wang	Modified complete case analysis	S(weighted least squares estimator (placed under weighting)	A weighted univariate analysis of generalized least squares (GLS)
418	Wang	Modified complete case analysis	The unweighted least squares (UWLS) estimator	A least squares estimator with no weights.
418	Wang	Modified mixed model	ANCOVA (The 'ANCOVA' estimator and the weighted mean estimator)	A weighted analysis of covariance estimator using an individual's LS slope as the outcome and their dropout time as the covariance. The linear minimum least squares estimator is used to estimate a mean LS slope for each treatment arm for each missing value. A weight based
418	Wang	Classic mixed model	The maximum ikelihood estimator (ML)	The maximum likelihood estimate under a normality assumption for the outcome. There is no requirement for a minimum number of
439	Hu	Modified robust regression	Semi-Paramateric (SEM)	The missing value is estimated using non-parametric function, where the dimension is reduced using parametric working index.
439	Hu	Modified robust regression	AIPW	No specific information or description provided
439	Hu	Classic robust regression	IPW	No specific information or description provided
496	Hedden	Classic mixed model	Mixed Effect Model (MEM)	It uses all the observed data for the analysis.

496	Hedden	Classic multiple imputation	Multiple Imputation (MI)	A regression model is fit for each outcome variable with missing data where the independent variables of the model are the previously observed outcome variables. Using the parameter estimates of the fitted model, a new regression model is simulated using posterior predictive distribution. This new model is used to impute the missing
496	Hedden	Modified complete case analysis	(Modified) stratified summary statistics (SSS)	Each individual measurement is reduced to a single summary measure. The summary statistics may include the area under the curve, time to peak response, rate of change over time, linear combinations, nonlinear functions, order statistics and survival functions. Modified is when a t test is used rather than z test for
496	Hedden	Classic LOCF	Last Observation Carried Forward (LOCF)	It is a single value imputation method where the last observation is carried forward until the end of the trial.
528	Liang	Classic single imputation	Empirical likelihood based confidence interval (EL)	The empirical likelihood method is used to combine the imputation for the missing data to construct a confidence interval.
528	Liang	Classic single imputation	LZ method (Liu 2006 method)	A mean-of-ratio imputation approach to impute the missing data for finite population.
528	Liang	Classic robust regression	RS method (Rao and Sitter, 1995 method)	It uses a ratio imputation approach to impute the missing values in a finite population
528	Liang	Classic LOCF	Modified last-value-carried-forward (MLVCF)	where a random noise was added when carrying forward the last observed value to correct the underestimation of variability due to
546	Barnes	Classic LOCF	Last Observation Carried Forward (LOCF)	The last non-missing observation is used as the end-point score for all subsequent missing
546	Barnes	Classic LOCF	Baseline Observation Carried Forward (BOCF)	Baseline observations are used to fill in any post-baseline missing data
546	Barnes	Classic multiple imputation	Multiple Imputation using PROC MI in Software SAS	Bayesian least square approach is implemented using SAS PROC MI and MIANALYZE. The model for imputation is identical to that of the
546	Barnes	Classic mixed model	Likelihood-based repeated measures (MMRM)	Parameters were estimated using restricted maximum likelihood with Newton-Raphson algorithm.
578	Baron	Classic multiple imputation	Last Observation Carried Forward (LOCF)	The last non-missing observation is used as the end-point score for all subsequent missing
578	Baron	Classic mixed model	Multipe Imputation using Markov chain Monte Carlo (MCMC)	Missing values are replaced using the Markov chain Monte Carlo replacement method
578	Baron	Complete case	Complete case analysis	Keep only individuals with complete data.
727	Mallinckrodt	Classic and modified mixed effects (methods analysed over both MNAR and MAR data)	Likelihood-based repeated measures (MMRM)	Parameter estimation was performed using Restricted Maximum Likelihood with the Newton-Raphson algorithm. All time periods were set as the dependent variable.
727	Mallinckrodt	Classic and modified LOCF (methods analysed over both MNAR and MAR data)	Last observation carried forward (LOCF)	LOCF with a fixed effects analysis of variance
737	Wei	Modified LOCF	Partial Imputation (PI)	The combination of last value carried forward and all available data approaches. This method partially carries forward the last
737	Wei	Classic LOCF	Last Value Carried Forward (LVCF)	The approach is identical to LOCF
737	Wei	Classic mixed model	Mixed Effect Model (MEM)	Used to model the covariance among the observations at different time points. Several configurations were considered, from simple to
737	Wei	Classic complete case analysis	All available data (AAD)	Uses all the available data for the analysis and doesn't impute missing data
897	Roderick	Classsic mixed model	Mixed-effect hybrid models (MEHM)	The joint distribution of the outcome and missing process is factorized into a marginal distribution of a random-effect model. The missing process is conditioned on random effect while the outcome process is
897	Roderick	Classic LOCF	Shared-parameter models (SPM)	One type of mixed-effect hybrid model where the same outcome process is assumed for different missing patterns. The missing process is implicitly correlated with the outcome process through modeling
897	Roderick	Classic mixed model	Random-effect model (REM)	The likelihood of the observed data is determined through integration of the missing data and random effects from the joint distribution of
1013	Guo	Classic mixed model	Random parameter mixture model	Mixed effects model that stratify the data according to time to dropout and form a model for each stratum
1013	Guo	Modified mixed model	Shared-parameter model	Assumes the longitudinal outcomes are directly linked to the true dropout times before censoring
1234	Carpenter	Classic multiple imputation	Multiple imputation using PROC MI	Applies the Markov Chain Monte Carlo (MCMC) method to impute missing values

1234	Carpenter	Modified multiple imputation	Re-weighting MI	Importance sampling is used to re-weight parameter estimates obtained from imputed data sets (via MCMC). An application of the
1274	Yuan	Modified mixed model	Random-effect Model	The first follows conventional estimation of the missing-not-at-random model
1274	Yuan	Modified mixed model	SC Model	The second reflects the concerns expressed above about the dependence of such an analysis on untestable assumptions, and consists of a sensitivity analysis derived from a series of model fits.
1569	Tsonaka	Modified mixed model	Semi-Parametric Shared Parameter model (SPSP)	Parametric assumptions for the random effects distribution are avoided and instead left unspecified. Model estimation is made using
1569	Tsonaka	Modified mixed model	Shared parameter model (SPM)	A framework for the joint modeling of measurement and missingness processes. A mixed model with a set of random effects containing parametric assumptions for their distribution.
1570	Cook	Classic single imputation	Expected value imputation	Estimates from the expected value of the missing observations conditional on the observed data are used for imputation.
1570	Cook	Modified multiple imputation	Multiple Imputation using expected value imptuation and EM algorithm	Perform expected value imputation to fill all missing values of dataset and then determine maximum likelihood estimates of the parameters using the 'complete' data. The process is repeated iteratively until the parameter estimates converge. PROC MI or BMDP 5V can be used in
1570	Cook	Classic LOCF	LOCF	The last available measurement is carried forward to fill in unobserved
1780	Gueorguieva	Modified mixed model	Joint models with dropout separated by reason	Joint model with competing risks
1780	Gueorguieva	Classic mixed model	Common dropout	A model in which all dropout was treated the same
1780	Gueorguieva	Classic mixed model	Ignoring dropout	A model for repeated measures outcome in which dropout was ignored
1821	Hogan	Modified complete	W1	Estimating function for the treatment difference delta_0 (assuming
1821	Hogan	Modified complete case analysis	W2	Estimating function of delta_0 assuming censoring variables are independent on the entire underlying processes
1821	Hogan	Modified complete case analysis	W3	An unbiased estimating function of delta_0

1868	Demirtas	Modified multiple imputation	RCPMM using a multivariate normal model with an unstructured co-variance matrix (NORM)	NORM uses a multivariate normal model with an unstructured covariance matrix and ignorable non-response; others correspond to conventional random-coefficient pattern-mixture models and incorporated with multiple imputation. A type of random-coefficient
1868	Demirtas	Modified multiple imputation	Random-coefficient pattern-mixture model (RCPMM) using a saturated polynomial model (FULL.POLY)	FULL.POLY's imputation is built on a saturated polynomial model that includes all estimable interactions of R and T and incorporated with multiple imputation. A type of random-coefficient pattern-mixture model. MCMC routines is used to create multiple imputations.
1868	Demirtas	Modified multiple imputation	RCPMM using a reduced polynomial model (RED.POLY)	RED.POLY uses trimmed version of full polynomials by a series of Wald's tests and incorporated with multiple imputation. A type of random-coefficient pattern-mixture model. MCMC routines is used to
1868	Demirtas	Modified multiple imputation	RCPMM using a Polynomial model (RLIN*TLIN)	RLIN*TLIN uses a polynomial model that includes only T, R and RT)RLIN×TLIN allows RT as the highest order term and incorporated with multiple imputation. A type of random-coefficient pattern-
1868	Demirtas	Modified multiple imputation	RCPMM using pattern-specific polynomial extrapolation (PATT.SPEC)	PATT.SPEC performs extrapolation within each drop-out pattern and incorporated with multiple imputation. A type of random-coefficient pattern-mixture model. MCMC routines is used to create multiple
1868	Demirtas	Modified multiple imputation	RCPMM using complete-case polynomial-coefficient restrictions (CCPC)	CCPC impose identifying restrictions over drop-out patterns by borrowing information from complete cases; incorporated with multiple imputation. A type of random-coefficient pattern-mixture
1868	Demirtas	Modified multiple imputation	RCPMM using neighbouring-case polynomial-coefficient-restrictions (NCPC)	NCPC impose identifying restrictions over drop-out patterns by borrowing information from neighbouring cases and incorporated with multiple imputation. A type of random-coefficient pattern-
1868	Demirtas	Modified multiple imputation	RCPMM using available-case polynomial-coefficient restrictions (ACPC)	ACPC impose identifying restrictions over drop-out patterns by borrowing information from available cases and incorporated with multiple imputation. MCMC routines is used to create multiple
1869	Demirtas	Modified multiple imputation	PMM-LIN	Conventional pattern mixture model, with a linear term for time of last measurement and its interactions with drug indicator and weeks of measurement. MCMC routines is used to create multiple
1869	Demirtas	Modified multiple imputation	PMM-DROP	Conventional pattern-mixture model with a dummy indicator of completion status and its interactions with drug indicator and weeks of measurement. MCMC routines is used to create multiple
1869	Demirtas	Modified multiple imputation	Hierarchical Bayesian pattern-mixture model (HBPMM)	A hierarchical Bayesian pattern-mixture model that allows decreasing amount of variability among high order time polynomials. MCMC routines is used to create multiple imputations.

1869	Demirtas	Classic LOCF	LOCF	Same as above definition
1869	Demirtas	Modified mixed model	complete case analysis	Same as above definition
1869	Demirtas	Modified mixed model	OSWALD-IGN using OSWALD software package	Assumes dropouts are ignorabeand dropouts depend on the wave of the measurements and the previous responses. A linear mixed model
1869	Demirtas	Modified mixed model	OSWALD-NIGN using OSWALD software package	Assumes dropouts depend on the wave of the measurement and the current response. A linear mixed model is used.
1870	Demirtas	Modified multiple imputation	Random-coefficient pattern-mixture model (RCPMM)	Random-coefficient pattern-mixture models: a marginal distribution for the time of drop-out combined with conditional distribution for the complete data given time of dropout. Different methods are used
1870	Demirtas	Modified mixed model	Ignorable	Ignorable, a random-coeffcient model with effects for G, T and GT with intercepts and time-slopes varying by patient
1870	Demirtas	Classic mixed model	Pattern-mixture model proposed by Hedeker and Gibbons (HedGib)	Hedgib, the pattern-mixture model used by Hedeker and Gibbons.
1870	Demirtas	Modified multiple imputation	Random-coefficient pattern-mixture model (RCPMM) using a saturated polynomial model (FULL.POLY)	FULL.POLY's imputation is built on a saturated polynomial model that includes all estimable interactions of R and T and incorporated with multiple imputation. A type of random-coefficient pattern-mixture model.
1870	Demirtas	Modified multiple imputation	RCPMM using a reduced polynomial model (RED.POLY)	RED.POLY uses trimmed version of full polynomials by a series of Wald's tests and incorporated with multiple imputation. A type of random-coefficient pattern-mixture model.
1870	Demirtas	Modified multiple imputation	RCPMM using a Polynomial model (RLIN*TLIN)	RLIN*TLIN uses a polynomial model that includes only T, R and RT)RLIN×TLIN allows RT as the highest order term and incorporated with multiple imputation. A type of random-coefficient pattern-
1870	Demirtas	Modified multiple imputation	RCPMM using pattern-specific polynomial extrapolation (PATT.SPEC)	PATT.SPEC performs extrapolation within each drop-out pattern and incorporated with multiple imputation. A type of random-coefficient pattern-mixture model.
1870	Demirtas	Modified multiple imputation	RCPMM using complete-case polynomial-coefficient restrictions (CCPC)	CCPC impose identifying restrictions over drop-out patterns by borrowing information from complete cases; incorporated with multiple imputation. A type of random-coefficient pattern-mixture
1870	Demirtas	Modified multiple imputation	RCPMM using neighbouring-case polynomial-coefficient-restrictions (NCPC)	NCPC impose identifying restrictions over drop-out patterns by borrowing information from neighbouring cases and incorporated with multiple imputation. A type of random-coefficient pattern-
1870	Demirtas	Modified multiple imputation	RCPMM using available-case polynomial-coefficient restrictions (ACPC)	ACPC impose identifying restrictions over drop-out patterns by borrowing information from available cases and incorporated with multiple imputation.
1870	Demirtas	Modified multiple imputation	RCPMM using a multivariate normal model with an unstructured co-variance matrix (NORM)	NORM uses a multivariate normal model with an unstructured covariance matrix and ignorable non-response; others correspond to conventional random-coefficient pattern-mixture models and incorporated with multiple imputation. A type of random-coefficient
2078	Longford	Classic LOCF	Bring last value forward (BLVF)	No specific information provided
2078	Longford	Classic multiple imputation	Multiple Imputation	Referred to as the stochastic version of EM algorithm. Values are drawn from a plausible estimated joint distribution of the missing values. The generate these plausible values, EM algorithm is used to
2078	Longford	Classic single imputation	EM Algorithm	Method used to find the maximum-likelihood estimate of the parameters when a data set has missing values
2078	Longford	Classic single imputation	Linear regression	statistical method used for estimating the relationships among variables
2215	Michael	Classic robust regression	Posterior predictive criterion	Used for model selection for incomplete longitudinal data.
2215	Michael	Classic robust regression	Selection Model (SM)	Used for model selection for bayesian inference with incomplete data.
2215	Michael	Modified mixed model	Mixture model 1 (MM1)	The most complex mixture model used for model selection.
2215	Michael	Classic mixed model	Mixture model 2 (MM2)	Mixture model used for model selection, which allows some equality of parameters between treatments.
2634	Gadbury	Classic multiple imputation	Mixed effect model	Model is based on maximum likelihood inference. It is also referred to as mixed linear models, two-stage random effects models, or random
2634	Gadbury	Classic mixed model	Multiple imputation	Same as above definition
2634	Gadbury	Classic LOCF	LOCF	Same as above definition
2634	Gadbury	Classic complete case analysis	Complete case analysis	Same as above definition
2799	Gilbert	Classic complete case analysis	Before deletion (BD)	Estimate causal effects by applying GBH/JR to complete data before deletion
2799	Gilbert	Classic complete case analysis	Complete case analysis (CC)	A complete case analysis that applies GBH/JR

2799	Gilbert	Classic complete case analysis	Ordinary least squares complete case (OLSCC)	Compares sample averages of outcome (Y3) between groups in all infected subjects that it measured.
2799	Gilbert	Modified robust regression	Spline propensity prediction method (SPPL)	A penalized spline propensity prediction based on a regression of Y3 on the spline of Y1∗, where Y1∗ is the linear predictor of the estimated propensity to observe Y3 from a linear logistic regression of
2816	Joseph	Classic mixture model	Varying Coefficients Model (VCM)	A mixture model for joint distribution for longitudinal repeated measures, where the dropout distribution may be continuous and the dependence between response and dropout is semi-parametric. Specifically, the assumptions were that responses follow a varying coefficient random effects model conditional on dropout time, where the regression coefficients depend on dropout time through unspecified nonparametric functions that are estimated using step
2816	Joseph	Classic mixture model	Conditional Linear Models(CLM)	Is a specialized version of the VCM with more stable estimates. Regression coefficients such as intercepts and slopes depend on
3093	Groenwold	Classic complete case analysis	Complete case analysis	Same as above definition
3093	Groenwold	Classic single imputation	Conditional mean imputation	For continuous outcome data, the most likely value based on the multivariable model can be imputed for the missing observation
3093	Groenwold	Modified multiple imputation	Multiple imputation with Regression (with random residual added to each predicted value)	Multiple imputation with missing data imputed with predicted values from a multivariable (regression) model that includes these observed patient data.
3133	James	Classic GEE	GEE (optimal)	Standard GEE, no definition provided.
3133	James	Modified GEE	SWEEP	A weighted model based data augmentation method which use the normal theory maximum likelihood estimator from Little & Rubin
3133	James	Modified GEE	Extended Sweep	A weighted model based data augmentation method which non-linear generalization of the sweep estimator
3133	James	Modified GEE	Extended GEE (optimal)	Defintion unclear
3133	James	Unclear	Nonmonotone	Defintion unclear
3369	Hebert	Classic LOCF	LDCF (same as LOCF)	The last non-missing observation is used as the end-point score for all
3369	Hebert	Classic LOCF	Next and Last observation (NAL)	The average of the last observed value (before the missing value) and the next observed value. If no next value exists, then the last non-
3369	Hebert	Classic single imputation	Random regression model (XB)	Predicted values determined from a regression model. Additional variance is added in the form of a mean zero normal disturbance with variance equal to the estimated variance of the regression error.
3369	Hebert	Classic single imputation	Single imputation regression with external data sources (XB+Cht):	Similar to a random regression model, except it includes the fitted values from a linear mixed model as an additional covariate. The is a function of time and time squared, with patient specific random
3369	Hebert	Classic multiple imputation	Multiple imputation	Multiple imputations based on the regression model used to predict XB
3369	Hebert	Classic multiple imputation	Multiple imputation (MI+Cht)	Imputations based on the regression model used to predict XB+ Cht. Inclusive of the chart-based fitted values as a covariate
3386	Lin	Classic GEE	5a-GEE	Doesn't require an explanation
3386	Lin	Modified GEE	5b-Visit intensity weighted GEE	A class of inverse intensity-of-visit process-weighted estimators in marginal regression models for longitudinal responses that may be
3386	Lin	Modified GEE	5c-Visit intensity weighted GEE	A class of inverse intensity-of-visit process-weighted estimators in marginal regression models for longitudinal responses that may be
3386	Lin	Modified GEE	5d-Visit intensity weighted GEE	A class of inverse intensity-of-visit process-weighted estimators in marginal regression models for longitudinal responses that may be
3438	Yi	Modified GEE	mean regression	The use of 'weighted' estimating equations in quantile regression models for longitudinal data with drop-outs is described. In particular, the conventional estimating equations for the quantile regression parameters are weighted inversely proportionally to the probability of
3438	Yi	Modified GEE	median regression	Median regression models to deal with longitudinal data with dropouts. Weighted estimating equations are proposed to estimate the median regression parameters for incomplete longitudinal data, where the weights are determined by modeling the dropout process. Median regression parameters used for incomplete longitudinal data.
3455	Stuart	Classic mixed model	Random effect model (RE)	In RE models, population (or group-specific) mean rate of change in the disease marker is a weighted mean of the subject-specific
3455	Stuart	Modified mixed model	Join multivariate random effect model (JMRE)	The method combines a linear random effects model for marker trajectory with a log-normal survival model for the informative
3742	Huson	Classic LOCF	LOCF	No specific information provided
3742	Huson	Classic single imputation	Baseline-carried-forward	The change from baseline to end point was set equal to zero.This type of conservative imputation method is often used in analysis of clinical trial data and such an approach is often referred to as a "worst-case"
3742	Huson	Classic single imputation	Nearest-neighbour hot-deck imputation	Adoption of future value(s) from a patient whose variable level is closest to the value the patient had at the time of drop-out, as long as the patient from which the values are being copied is not dropping out

3742	Huson	Modified LOCF	LOCF values analyzed using censored regression	the LOCF approach assumes that values for a patient post withdrawal would have remained constant. An obvious alternative assumption is that the values for a given patient after the time of withdrawal would have followed a trend had the patient remained in the study. This assumption is equivalent to regarding the value carried forward from the point of withdrawal as having been censored, and, therefore, an
3742	Huson	Classic multiple imputation	Multiple imputation	Monte-Carlo Markov Chain (MCMC) method for multiple imputation
4013	Wei	Modified robust regression	Simple rank estimation	In this article we propose a semiparametric estimation procedure for treatment differences over time based on repeated measurements of an outcome variable when the patient's follow-up time may depend
4013	Wei	Modified robust regression	Conditional linear model	A method explored by Wu and Bailey (1989) with the mean structure of the response variable follows a a growth curve model
4013	Wei	Classic robust regression	Naive Method	Treats censoring as non-informative; no definition in paper except in mathematical terms.
4502	James	Modified GEE	Inverse probability of censoring weighted (IPCW) estimating equations (3 types)	Three types of augmented inverse probability of response weighted estimators are described using simple and complex estimators referred to as NI(I) and NI(S), respetively. Full specification of a parametric likelihood is not needed for the estimators.
4502	James	Modified GEE	NI (I)	Please see above
4502	James	Modified GEE	NI (S)	Please see above
4604	Lane	Classic LOCF	LOCF	No specific information provided
4604	Lane	Classsic mixed model	Mixed model for repeated measurements (MMRM)	A likelihood-based approach that models jointly all the actual observations in a longitudinal trial, with no attempt at imputation or adjustment for the missing-value mechanism. The series of observations from an individual subject are treated as arising from a multivariate normal distribution, whose covariance matrix describes
4860	Li	Classic mixed model	Naïve method	Treats censored data as exact values. For right censored measures the upper limit is often used as the exact value. For measures that are interval censored, the lower bound is often used as the exact value. For measures that are left-censored at a specific limit (d_1), the values
4860	Li	Classic multiple imputation	Multiple imputation with regression model	Multiple imputation can be readily applied to estimate the population mean and standard deviation for censored data from the serial
4860	Li	Classic multiple imputation	Maximum likelihood	No specific information provided
4952	Liu	Classic multiple imputation	Joint mixed model (JMM)	Likelihood methods based on joint modeling
4952	Liu	Classic mixed model	Joint multiple imputation (JMI)	MI based on the joint modeling
4952	Liu	Modified mixed model	Separate mixed model (SMM)	No explanation provided (It appears to be the classic mixed model)
4952	Liu	Modified multiple imputation	Separate multiple imputation (SMI)	No explanation provided (It appears to be the classic multiple imputation)
4952	Liu	Classic LOCF	LVCF	The last non-missing observation is used as the end-point score for all
5464	Mehrotra	Classic mixed	restricted maximum	The parametric linear mixed effects model approach, with estimation
5464	Mehrotra	Classic GEE	Generalized estimating	GEE approach rests on the restrictive assumption that missing data
5464	Mehrotra	Modified GEE	Weighted extension of GEE (WGEE)	that is valid even when the missing data are MAR, provided that appropriate weights are used. Application of WGEE for trials with monotone patterns of missing data entails a conceptually simple two-step process. The first step requires fitting a model, usually a logistic regression model, for the probability of being observed (i.e., not missing) at each time point as a function of observed responses before that time point, relevant covariates, and/or auxiliary variables. In the second step, weighted estimating equations are used for estimation
5464	Mehrotra	Modified multiple imputation	Multiple imputation using PROC MI	Multiple imputation with robust regression analysis via M-estimation
6828	Revicicki	Classic single imputation	Arbitrary substitution	Assumption is made that death is an indicator of worsening health. Missing scores are assigned the previous complete assessment minus a decrement to account for the fact that the current assessment is
6828	Revicicki	Modified single imputation	Emprical bayes	The empiracl Bayes with informed censoring imputation method. Categorizes individuals with n-r sequential observations and r sequential missing observations due to mortality. A slope is estimated for each individual that is a weighted average of the individual's slope for available time points and the OLS estimate for the population
6828	Revicicki	Classic LOCF	LOCF	The last non-missing observation is used as the end-point score for all subsequent missing observations.
6828	Revicicki	Classic single imputation	Random regression	No specific information provided

6828	Revicicki	Classic single imputation	Within-Subject Modeling (WSMOD)	Missing values for outcomes for individuals is estimated on the basis of the nonmissing observations using ordinary least squares (OLS) regression. This requires observed data for a minimum of 2 measurement occasions and estimates the missing data based on the
7236	Scharfstein	Modified robust regression	Simple Inverse Weighted Estimator	In this estimator, the observed values of the outcomes are inverse weighted by the estimated probability of being observed. The denominator converges in probability to one. It can be shown that the resulting estimator is consistent and asymptotically normal, provided
7236	Scharfstein	Modified robust regression	Doubly Robust Estimator	Enough lower-dimensional restrictions (parametric or semiparametric) on the conditional distribution of Y given X and R=1 to ensure the ratio of expectations can be well estimated in finite samples. This estimator is called doubly robust, since it will be consistent and asymptotically normal (CAN) if model (2) or the lower-dimensional
7236	Scharfstein	Modified robust regression	Orthogonal Estimating Function Estimator	Used when a parameter of interest is orthogonal to the nuisance tangent space for another paremeter.
7543	Siddiqui	Classic multiple imputation	Multiple Imputation	Markov Chain Monte Carlo (MCMC) and regression methods were used for imputing data
7543	Siddiqui	Classic mixed model	Mixed-effects model repeated measures (MMRM)	A special form of the general mixed-effects regression model analysis. Time is set as a factor variable. The interaction (Treatment × Time) effect is set as an unstructured interaction effect, instead of considering Treatment ×Time effect as the slope (rate of change)
8006	Tang	Classic single imputation	Hot Deck	A modified predicted mean matching method is used for item-level missing data and an approximate Bayesian bootstrap (ABB) for unit-
8006	Tang	Classic multiple imputation	Multiple imputation based on an Multivariate Normal (MVN) model	Multiple imputation using a multivariate normal model.
8006	Tang	Classic LOCF	LOCF	No clear descrption provided
8006	Tang	Classic multiple imputation	Hot Deck multiple imputation	Combination of hot-deck (HD) multiple imputation using a predictive mean matching method for item non-response and the approximate
8006	Tang	Classic complete	Available-case (AC) analysis	No clear descrption provided
8215	Touloumi	Classic mixed model	Random effects (RE) model	The proposed method combines a linear random effects (RE) model for the underlying pattern of the marker with a log-normal survival model for the drop-out process. Estimators are obtained through
8215	Touloumi	Modified mixed model	joint multivariate random effects (MRE) model	A nested EM algorithm is applied to incorporate survival-censored data. The superscript (joint) denotes the joint multivariate random
8332	Unnebrink	Classic complete case analysis	CCA	Keep and analyze only individuals with complete data.
8332	Unnebrink	Classic LOCF	LOCF	The probably most frequently used approach for dealing with missing values in clinical trials with continuous variables is to use the last value obtained from a patient for the analysis, called 'endpoint analysis' or 'last observation carried forward' (LOCF). This strategy includes data of a patient's last known state in the analysis, implicitly assuming that
8332	Unnebrink	Classic multiple imputation	mean of the missing patient's own group (MOWN)	The mean of the missing patient's own group (MOWN) assumes that this patient is representative for his group and not systematically different from those patients fully observed. Impute the group
8332	Unnebrink	Classic multiple imputation	mean of the other group (MOTH)	Imputation of the mean of the other group (MOTH)
8332	Unnebrink	Classic single imputation	RegOWN	Regression based on observed patients of own group
8332	Unnebrink	Classic single imputation	RegOTH	Regression based on observed patients of other group
8332	Unnebrink	Classic single imputation	RegP	Regression based on the observed patients of the placebo group (RegP) implicitly assumes that treatment is stopped at the same time as observation and that therefore the patient from then on continues
8332	Unnebrink	Classic single imputation	RegM	Minimax-regression (RegM) is a strategy ensuring the conservativity of the results. Missing data for patients in the placebo group are imputed by the worst, in the treatment group by the best mean yearly increase
8332	Unnebrink	Classic single imputation	RankG	A ranking strategy (RankG), ranking drop-outs on the worst rank behind all observed patients. If, however, patients definitely drop out because of cure, they are to be ranked in front of the observed
8332	Unnebrink	Classic single imputation	RankS	Consists of dividing the worst ranks between the patients with missing values depending on their time of drop-out. The earlier the time of
8332	Unnebrink	Classic single imputation	Dicho	It was investigated dichotomization of the continuous endpoint, counting drop-outs as treatment failures. The cutpoint for

8332	Unnebrink	Classic single imputation	Worst and best case analysis	The results for worst case and best-case analysis were reported. In worst case analysis (Worst), the worst scenario for the treatment group is assumed, that is, in the treatment group drop-outs occur only due to failure (inclusion in the analysis with the worst rank possible), in the control group only due to success or cure (inclusion with the best rank). Contrarily, in best case analysis (Best) the best scenario for
8952	Wu	Modified robust regression	Wu and Bailey	An approximate conditional linear approach with fewer assumptions on the drop-out probability mechanism. It uses a second stage regression of the random effects on summary measures of the missingness (e.g. time of drop-out, total number of missed visits, etc). A type of linear mixed model in which the summary measures are
8952	Wu	Modified robust regression	Probit time-independent informative drop-out process	When drop-out is informative, the probability of drop-out can be dependent on the initial value and slope of the primary response. This dependency remains constant over time. It is referred to as time-
8952	Wu	Modified robust regression	Probit Time-Independent informative drop-out process	Logistic time-dependent informative drop-out model.
8963	Xu	Classic complete case analysis	Complete Case analysis	No additional information provided
8963	Xu	Classic complete case analysis	Available Case Analysis	No additional information provided
8963	Xu	Classic single imputation	Intent-to-treat analysis	Missing is assumed to imply no change.
8963	Xu	Classic complete case analysis	Mean-imputed analysis (imputing missing values with group means)	Mean comparison for matched-pair samples with missing data. New test statistics are proposed to use all available information with asymptotic expansion for the null distributions under general
9037	Yuan	Classic all available data	before deletion	Quantile regression is applied to complete data before deletion
9037	Yuan	Classic robust regression	random quantile regression	Quantile regression with a shared latent individual-specific random effect. Quantile regression models the lower or higher quantiles of the outcome which allow for a more natural analysis of covariate effects specific for those regression quantiles. Bayesian MCMC estimation method is used. Application can be found in longitudinal
9037	Yuan	Modified robust regression	shared parameter quantile regression	Individual-level quantile regression (QR) parameters are shrunk toward a population value by penalizing the standard check function of QR. Quantile regression models the lower or higher quantiles of the outcome which allow for a more natural analysis of covariate effects specific for those regression quantiles. Bayesian MCMC
9473	Tseng	Classic mixed model	Linear mixed effects model	Model linear mixed effects model
9473	Tseng	Modified robust regression	T model	Maximum pseudo-likelihood estimation based on t distribution model allows NMAR
9473	Tseng	Modified robust regression	Robust normal model	Maximum pseudo-likelihood estimation based on robust normal likelihood
9473	Tseng	Classic mixed model	Normal model	Maximum pseudo-likelihood estimation based on normal distribution model

CHAPTER 4:

REPORTING AND METHODS FOR HANDLING MISSING PARTICIPANT DATA FOR CNTINUOUS OUTCOMES IN RANDOMIZED CONTROLLED TRIALS: A SYSTEMATIC SURVEY

Yuqing Zhang [1], Ivan D. Flórez[1,2], Luis E. Colunga Lozano[3], Fazila Abu Bakar Aloweni [4], Sean Alexander Kennedy [5], Aihua Li[1], Samantha Craigie[6], Shiyuan Zhang [7], Arnav Agarwal [8], Luciane C Lopes[9,10], Tahira Devji[1], Wojtek Wiercioch[1], John J. Riva[1,11] Mengxiao Wang[1], Xuejing Jin[1], Yutong Fei [12], Paul Alexander [1], Gian Paolo Morgano[1], Yuan Zhang[1], Alonso Carrasco-Labra[1,13], Lara A Kahale[14], David Meyre[1,15], Elie A. Akl[1,14], Holger J. Schünemann[17], Lehana Thabane[1], Gordon Guyatt[16]

[1] Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
[2] University of Antioquia, School of Medicine, Medellín, Colombia
[3] Hospital Civil de Guadalajara, "Fray Antonio Alcalde", México
[4] Singapore General Hospital, Singapore
[5] Department of Diagnostic Radiology, University of Toronto, Toronto, Ontario, Canada
[6] Michael G. DeGroote National Pain Centre, McMaster University, Canada
[7] Health Economics and Outcomes Research, Medical Affairs, GSK, Canada
[8] Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[9] Universidade de Sorocaba, São Paulo, Brasil
[10] Universidade Estadual Paulista "Julio de Mesquita Filho", São Paulo, Brazil.
[11] Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada
[12] Beijing University of Chinese Medicine, McMaster University, China, Canada
[13] Evidence-Based Dentistry Unit Faculty of Dentistry, Universidad de Chile Santiago, Chile
[14] Department of Internal Medicine, American University of Beirut, Lebanon
[15] Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada
[16] Department of Medicine and Clinical Epidemiology and Biostatistics, McMaster University, Canada

Abstract

Background: Missing participant data (MPD) can bias randomized controlled trials' (RCTs) results if it is associated with the outcome of interest. To the best of our knowledge, no study has assessed the reporting and analytic approaches to MPD for continuous outcomes in RCTs.

Objective: To assess (1) how trial authors report MPD for patient-important continuous outcomes, and (2) the analytic approaches for their primary and sensitivity analyses to address MPD.

Method: We conducted a systematic survey of a representative sample of RCTs published in 2014 in the core medical journals. Eligible RCTs reported at least one patient-important outcome analyzed as a continuous variable.

Result: Among 200 studies, 187 (93.5%) trials explicitly reported whether MPD occurred. In the 163 (81.5%) trials that reported the occurrence of MPD, the median and interquartile ranges of the percentage of participants with MPD were 11.4% (2.5-22.6%). It was unclear in 16 (16.5%) of these 163 trials how trialists dealt with MPD in their primary analysis. Linear regression showed an association between larger sample size and a higher percentage of MPD.

Among the remaining 147 trials, the approaches trialists used in the primary analysis for MPD, by descending order of frequency were: using only complete data (109 trials, 67%), mixed effect models (10, 6.1%), last observation carrying forward (9, 4.5%), multiple imputation (9, 4.5%), maximum likelihood (3, 1.8%), mean imputation (2, 1.2%), regression (1, 0.6%), or other methods (4, 2.5%). Of 163 studies reporting MPD, 16

(9.8%) conducted sensitivity analyses examining the impact of the MPD. Very few trials (18, 11.1%) discussed the risk of bias associated with MPD .

Conclusion: Randomized trials reporting continuous outcomes typically have over 10% of participant data missing, and over 15% of trials fail to make clear how they deal with MPD in their analysis. When authors did make this clear, most conducted complete case analysis without taking into account missing data, and very few conducted sensitivity analyses addressing the possible impact of MPD or comment on how MPD might influence risk of bias.

Introduction:

Missing participant data (MPD) in randomized controlled trials (RCTs) - also referred to loss to follow-up, discontinued prematurely, or outcome not assessable[1] - refers to missing information on outcomes of interest[2]. Analyzing patients in the groups to which they were randomized will avoid bias for patients with complete data[3-5]. It does not address bias due to MPD, which, if it is substantial and the reasons for MPD differ between the intervention and control groups, is likely to bias the results. For instance, if patients with poorer quality of life at study termination withdraw consent more frequently from the intervention group than from the control group, and were excluded from the analysis, the results could be biased in favour of the treatment.

A common classification of the reason for missing data (also called missing mechanism) includes missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR)[6]. When outcome data is MCAR, it indicates no systematic differences between missing and observed values. It has been argued that in those cases ignoring missing data and only analyzing those with available data (complete case analysis) will not bias estimates of effect. Outcome data MAR denotes an explainable systematic difference between missing and observed values based on observed data. Ignoring missing data may cause bias in this case and imputation or data augmentation methods may reduce the extent of bias. When outcome data is NMAR, systematic differences between missing and observed values can only be explained by unobserved data (e.g. a person not responding to treatment is more likely not to provide an

observation)[7]. NMAR requires conducting sensitivity analysis comparing effect estimates under different missing mechanisms[6,8]. Since there is no detailed guidance on how to judge the missing mechanism, seldom can investigators be confident that their data is MCAR; thus, assuming some degree of MAR or NMAR is likely to be a more appropriate approach.

Despite the fact that investigators often expend enormous effort to prevent MPD, across all therapeutic areas, up to 89% of the RCTs report dichotomous MPD[9-12]. Under plausible assumptions about the outcomes of participants with missing data, up to one third of RCTs claiming treatment effects published in leading medical journals may lose statistical significance[12]. Researchers have thoroughly investigated how RCT authors have dealt with MPD in studies focusing on dichotomous outcomes[1,12,13]. Comparing to dichotomous outcomes, dealing with continuous MPD has its special challenges such as different distributional assumptions for observed data and the applicability of statistical models[14]. Considering the serious threat of bias from MPD, statisticians and methodologists have developed a variety of methods to deal with MPD in RCTs focusing on continuous outcomes[15-20]. Whether trialists are planning and applying the optimal approaches to handle continuous MPD is unknown.

Following our previous investigation addressing dichotomous MPD (LOST-IT)[12], the current study aims to assess 1) how trial authors report MPD for patient-important continuous outcomes, and 2) the analytic approaches for their primary and sensitivity analyses to address MPD.

Methods:

Definitions

We defined MPD as unavailable data from trial participants that, if available, would have been included in the analysis of the specific outcome in RCTs. We defined a patient-important outcome as an outcome for which a patient would say "yes" to the following question: "If this outcome were the only thing to change with treatment, would the patient consider receiving this treatment if it is associated with burden, side effects or cost?"[13]. We used a taxonomy characterizing a hierarchy of the importance of outcomes to select one outcome of primary interest from each trial (Appendix 1). Patient-important continuous outcomes high on this hierarchy include quality of life, symptoms and functional status. We did not consider surrogate outcomes as patient-important outcomes.

We defined complete case analysis as excluding all patients with missing value for the outcome being analyzed[21]. In contrast to the complete case analysis, all available data analysis refers to using all available observations for a particular outcome; this means including data from patients even with some missing values for that outcome. All available data analysis is commonly seen in trials with repeated measures[2].

Eligibility criteria

Inclusion criteria:

Eligible studies fulfilled all of the following criteria:

- Published in 2014 in one of 119 core **Abridged Index Medicus Journal Titles**

clinical journals, also known as the core medical journals;

- Described by authors as a RCT;
- Reported an analysis of data for at least one patient-important outcome analyzed as a continuous variable

Exclusion criteria:

We excluded studies meeting any of the following criteria:

- RCT reporting time to event **outcomes and analyze those as continuous data;**
- Non-human trials;
- Cluster RCT, factorial RCT, cross-over RCT, n-of-1 trials, cost-utility studies;
- Studies reporting continuous outcomes but analyzed as dichotomous data;
- Meta-analysis of two or more previously published RCTs;
- Secondary analysis of RCTs.

Literature search

We conducted the search using the Cochrane Collaboration's highly sensitive search strategy to identify RCTs through Medline (OVID interface) in the 119 core clinical English journals indexed under Abridged Index Medicus by the National Library of Medicine (available at http://www.nlm.nih.gov/bsd/aim.html)(see appendix 2).

Random sampling of citations

We retrieved a random sample of the identified citations using generated random numbers from an Excel sheet and retrieved correspondingly citation numbers. We repeatedly

sampled and screened identified citations meeting eligibility criteria until we achieved the target sample size.

Study selection and data collection

A team of 20 reviewers, with health research methodology training, worked in pairs using standardized forms to conduct screening of title and abstract, screening of full text, and data abstraction, all independently and in duplicate. We applied a calibration process prior to screening and data abstraction to ensure accuracy. Regarding screening and data abstraction process, reviewers resolved disagreement through discussion and with the assistance from an independent arbitrator (YZ), if needed. We also reviewed supplementary documents published by the authors to abstract information on detailed description on the reporting and analysis of MPD when authors referred to it. We conducted screening and data abstraction using a web-based systematic review software (DistillerSRTM; https://systematic-review.ca).

Selection of outcome and comparison

For RCTs including more than one patient-important continuous outcome, we selected the primary outcome as the authors reported. If authors reported more than one primary continuous outcome, we selected the first one reported in the abstract. For RCTs including more than one patient-important continuous outcome with none reported as the primary outcome, we selected the outcome first reported in the abstract, or in the results if not presented in the abstract.

In multiple-arm RCTs, we considered the first comparison reported in the results. For RCTs with multiple follow-up times, we used the analysis that included all time points or, if there was no such analysis, the analysis focused on the longest follow-up time.

Data abstraction

For each trial, we abstracted data regarding general characteristics, methodological characteristics, reporting, planned and conducted analytic approach regarding MPD, and the extent of MPD. We recorded the categories trial investigators used to describe participants with potential MPD including: ineligible participants, mistakenly randomized, did not receive any intervention, withdrew consent, dead, experienced adverse events, non-compliant or non-adherent, discontinued prematurely, excluded as part of center exclusion, and outcome not assessable. We also recorded the missing mechanism the trial assumed when dealing with MPD, whether authors reported a justification of their approach to MPD, as well as whether trialist assessed risk of bias associated with MPD.

Sample size:

We chose a sample size to achieve a precise confidence interval (+/-0.05) around the proportion of RCTs that conducted primary analytical approach regarding MPD. In the most conservative situation in which the proportion is 0.5, we would need 200 RCTs to achieve the desired confidence interval (0.45, 0.55).

Analysis

We assessed agreement for eligibility between reviewers at both the title and abstract screening stage and the full text screening using kappa statistics. We followed the interpretation guideline from Landis and Koch[22]: kappa values of 0 to 0.20 represent slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and greater than 0.80 almost perfect agreement.

For all descriptive analyses, we used absolute number and percentages for dichotomous (categorical) variables and mean with standard deviation for continuous outcomes when distribution was normal or near normal. When the distribution was skewed to a large extent, we used median and interquartile range (IQR).

Categories of trial participants investigators considered as having MPD

For all the categories that trial investigators used to describe participants with potential MPD, such as "ineligible participants", "withdrew consent", "outcome not assessable", we reported the number and percentage of trials documenting the categories.

Reporting and extent of MPD

We calculated the percentage of participants with MPD in each trial and the median and interquartile range of the percentage across all trials. For trials with multiple follow-up times, in addition to these analyses we also calculated:

- The percentage of missing data points overall through the entire follow-up counted as the number of missing data points divided by total number of possible data points.

- At the last follow-up time, the percentage of missing data points counted as the number of missing data points divided by the total number of possible data points.

We planned to conduct a logistic regression in which the dependent variable was whether trials did or did not report MPD and the independent variables were:

- Sample size
- Type of intervention (pharmaceutical *vs.* surgical/invasive non surgical *vs.* others)
- Type of funding (for profit *vs.* not for profit *vs.* no funding reported)
- Journal type (top 5 *vs.* non-top 5)

  Top 5 refers to the five general medical journals with the highest impact factor in 2015: *Annals of Internal Medicine*, *British Medical Journal*, *Journal of the American Medical Association*, *The Lancet*, and *New England Journal of Medicine (http://impactfactor.weebly.com/medicine.html)*.
- Allocation concealment (inadequate *vs.* adequate)

Our *a priori* hypotheses were: trials with smaller sample size, for-profit type of funding, non-pharmaceutical type of intervention, inadequate allocation concealment and non-top 5 medical journal were less likely to report MPD.

We also conducted a linear regression with "the percentage of MPD" as dependent variable and the same independent variables described above. Our *a priori* hypotheses were: trials with larger sample size, for-profit type of funding, non-pharmaceutical type of intervention, inadequate allocation concealment and non-top 5 medical journal tended to have higher percentage of participants with MPD.

Planning and conduct of analyses addressing MPD

We conducted a descriptive analysis of the planned analysis regarding MPD for all continuous outcomes and the analysis conducted by trial investigators for the chosen outcome. We planned to conduct a logistic regression with whether trials planned a sensitivity analysis regarding MPD as the dependant variable and the independent variables in noted above. Our *a priori* hypotheses were: trials with smaller sample size, for-profit type of funding, non-pharmaceutical type of intervention, inadequate allocation concealment, and non-top 5 medical journal would be less likely to plan a sensitivity analysis regarding MPD.

The analysis was performed using the SPSS software, version 22/12 (IBM Corp, Tx).

Results

General characteristics of included RCTs

We included 200 eligible trials that met out target sample size (figure 1). Agreement between reviewers was substantial: kappa of 0.63 for title and abstract screening and 0.64 for full text screening.

Table 1 presents the general trial characteristics and Table 2 the methodological characteristics of the included studies. Symptoms (84, 42%), quality of life (44, 22%), and functional status (33, 16.5%) were the most frequently investigated continuous patient-important outcomes. All but one trials that reported time at which patients were followed up, the median follow-up time was 3.3 months (interquartile range of 0.7 to 12

months). Of these 199 trials, 92 (46%) reported a single follow-up time and 107 (53.5%) multiple follow-up times.

Reporting and extent of MPD

Table 3 presents information regarding the reporting of missing participant data. Among all 200 included trials, 187 (93.5%) had, in the main text or CONSORT flow diagram, an explicit statement of whether MPD occurred. Among the 187 trials that explicitly reported the presence or absence of MPD, 24 (12%) studies explicitly stated MPD did not occur, and 163 (81.5%) explicitly reported the extent of MPD, of which 44 (27%) trials reported the percentage of MPD in each arm and overall; the overall median and interquartile range of participants in all time points with MPD were 11.4% (2.5-22.6%). The reporting of MPD was mainly focused on the number of patients who had MPD for the overall study sample but not by the specific outcome.

For 91 trials that included multiple follow-up times and reported MPD from either overall or per arm or both, the medians and interquartile ranges for the percentage of total missing data points were 13.1% (6.1-23.7%). At the last follow-up time, the medians and interquartile ranges for the percentage of missing data points were 14.4% (7.4-23.6%). None of the differences between intervention and control in the frequency of missing data approached conventional levels of statistical significance.

We could not conduct the logistic regression with the dependent variable explicitly reporting (or not) the occurrence of MPD because of the small number of studies (13, 6.5%) that failed to report whether MPD occurred.

We conducted a multiple linear regression addressing the percentage of participants with MPD based on sample size, type of intervention, funding, journal, and allocation concealment. A significant beta coefficient indicated that there was a higher percentage MPD when sample size was larger (Beta coefficient 0.01(0.0-0.02), p=0.005, meaning the MPD would be 1% more for each 100 patients), with an R^2 of 0.167 (Appendix 3). We further explored the correlation between larger sample size and higher percentage of missing data using a bivariate analysis and found a correlation coefficient of 0.29 (p=0.001). Another significant beta coefficient indicated that there was a lower percentage of MPD when funding was not explicitly reported (Beta coefficient 4.89 (0.40-9.38), p=0.03). There was a trend that a larger percentage of participants with MPD may be associated with inadequate allocation concealment (beta coefficient -6.52(-13.14-0.11), p=0.05).

Categories of trial participants trial investigators considered as having MPD

Table 4 provides data regarding studies' reports of the reasons for MPD. Of 200 included studies, 24 explicitly reported absence of MPD, the remaining 176 studies potentially had MPD. The most frequently considered categories for potential MPD were "withdrew consent" (81 trials, 46.0%) and "experienced adverse event" (41 studies, 23.3%).

Analyses planned regarding MPD

Table 5 presents the analysis plan reported in the methods section of included trials. Among all 200 included studies, 58 (29%) and 21 (10.5%) reported, in the methods section of their article, a plan to handle MPD in their primary and sensitivity analysis,

respectively. The most frequent approaches specified were last outcome carried forward (LOCF) (11, 5.5%) and mixed effect model (11, 5.5%) for the primary analysis, multiple imputation (MI) for the sensitivity analysis regarding MPD.

Analyses conducted regarding MPD

Table 6 presents the analysis approaches authors used regarding MPD. Among 163 trials explicitly reporting the occurrence of MPD, it was unclear how trialists dealt with MPD in their primary analysis in 16 (9.8%) trials. Of the remaining 147 trials, 74 (45.4%) used complete case analysis, 35 (21.5%) all available data analysis, 9 (5.5%) LOCF, 10 (6.1%) mixed effect model, 9 (5.5%) MI, 3 (1.8%) maximum likelihood, 2 (1.2%) mean imputation, 1 (0.6%) regression, and 4 (2.5%) methods other than the above mentioned. Very few (14, 8.6%) trials specified the missing mechanism when conducting an analysis regarding MPD; in 13 of 14 that did make such an explicit statement, the assumption was MAR (table 5).

Regarding discordance between planned and conducted sensitivity analyses, among the 163 (81.5%) trials that explicitly reported occurrence of MPD, 138 (84.7%) trials neither reported in their methods a plan to conduct such analyses and did not do so, 11 (6.7%) trials reported a plan and reported results, 9 (5.5%) reported a plan but did not report results, and 5 (3.1%) did not report a plan in their methods but reported results of a conducted sensitivity analysis regarding MPD (table 7).

Of these 163 studies, 16 (9.8%) studies conducted sensitivity analysis for MPD (more than one sensitivity analysis can be conducted); multiple imputation (4, 2.5%), complete

case analysis (3, 1.8%), LOCF (2, 1.2%), mean imputation (1, 0.6%), combination of more than one method (1, 0.6%), other methods not mentioned above (2, 1.2%), and not reported (6, 4.5%). Of the 16 trials that reported results of a sensitivity analyses to assess the impact of MPD, 2 reported that results were no longer statistically significant in one of their sensitivity analyses. Of 163 trials reporting the occurrence of MPD, 18 (11.1%) discussed the implications of MPD regarding risk of bias.

We could not perform the planned logistic regression to explore the factors associated with whether trials planned sensitivity analysis regarding MPD, due to the small number of studies conducting sensitivity analysis.

Discussion

Summary of findings

Almost all trials 187 (93.5%) made explicit statements regarding MPD, of which 163 (81.5%) reported the occurrence and extent of MPD (table 2). Many of these 187 trials reported substantial MPD (median 11.4%, Q1 2.5% to Q3 22.6%).

Very few (14, 8.6%) trials specified the missing mechanism when conducting an analysis regarding MPD; in 13 of 14 that did make such an explicit statement, the assumption was MAR (table 6). The most common way trialists handled MPD was a complete case or all available data analysis (109, 67%). Other approaches included mixed effect models (10, 6.1%), LOCF (9, 4.5%), and MI (9, 4.5%).

Among all 200 trials, less than a third (58 trials, 29%) planned an analytical approach to

address MPD in their primary analysis and even fewer (21 trials, 10.5%) reported a plan for sensitivity analysis (table 5).

Strengths and limitations of study

Strengths of our study include a systematic and comprehensive search, independent and duplicate screening and data abstraction, and a focus on patient-important continuous outcomes. We also implemented standardized built-in instructions in both screening and data abstraction forms on the web-based systematic review software and conducted calibration exercises. Our random sample of eligible studies from the 119 core medical journals published in 2014 ensures high representativeness of the most updated practise among trialists and generalizability of our results[23-25].

One of the limitations is that we only captured the information authors reported in the publication and in the additional information provided in the appendix and supplementary data files. We might not have captured what authors have done beyond what they reported with respect to MPD. We could have contacted authors to clarify things that were unclear in the publications. The other limitation is we could have checked registered protocols of trials for their MPD analyses plan when applicable. Lastly, we should have adjusted for potential clustering effect for papers published in the same journal since they might have followed the same requirement from the journal to report the manuscript in a certain manner.

Comparing Findings with Other Studies

Akl *et al.*[12] investigated the extent of MPD, the reporting, and the impact on results associated with MPD in studies addressing binary outcomes in five general prestigious medical journals. They found 13% of the trials did not report whether MPD occurred and 20% did not clearly report the analytical approaches used to handle MPD. These results are very similar to what we found with respect to continuous outcomes. Alshurafa *et al.*[26] investigated how methodological articles defined intention to treat (ITT) analysis in the context of MPD. They found the most frequently mentioned strategies suggested to deal with MPD within ITT were LOCF (50%), general sensitivity analysis (50%), and imputation using all available data (46%). We found investigators took advantage (though infrequently) of a wider variety of sophisticated statistical strategies, and did not frequently use LOCF.

Interpretation of Findings

In a recently conducted systematic survey[27]we conducted on the performance of methods of handling continuous MPD, LOCF proved to be one of the worst methods among more than 14 categories and 250 methods investigated.

In our study, investigators seem aware of the limitations of LOCF, with only 9 of 200 studies using the method, which is likely much less than in the past. Also among 58 studies that reported analysis plan to deal with MPD in their primary analysis, LOCF is one of the most frequently chosen method reported in 11 studies. On the other hand, mixed effects models, which proved one of the best methods in the survey, was used or planned no more frequently than LOCF. Investigators can still make better choices in

choosing approaches to MPD.

We found an association between explicit reporting of MPD with explicit reporting of funding. This showed a potential association between not reporting MPD and poor reporting of other trial aspects. Not surprisingly, we also found trials with larger sample size had larger percentages of missing data. This finding highlights both the challenges of minimizing MPD in larger trials and enhances the importance of planning optimal analytical strategies to handle potential MPD.

Implications for trialists

Trial investigators should be more explicit in providing details on the reporting of MPD both at participant level and at outcome level. Particularly when trials have multiple follow-up times as is commonly seen in the context of continuous outcomes. Reporting only the number of patients missing without specification of MPD at outcome level may obscure the clarity of the extent of MPD.

Ideally, there are a number of strategies trial investigators may institute to minimize the risk of bias due to MPD. The first is to institute measures to minimize MPD[28]. MPD is often, however, inevitable, and strategies to deal effectively with MPD once it occurs will be necessary. These include developing in advance a plan to deal with MPD and reporting that plan in their protocol and ultimately in the methods section of their manuscripts. Investigators should determine if baseline characteristics and other covariates differ between patients with missing data and patients with complete data. Differences in characteristics suggest that data might be MAR or that they could even be

NMAR. Further, they should examine the relation between patients' characteristics and observed outcomes; if there are substantial associations it also suggests that to some extent the data are MAR and that imputation and data augmentation methods may be useful, at least as sensitivity analyses. The failure to find such associations suggests data might be NMAR and again mandates sensitivity analyses that include a wide range of plausible results for MPD.

Investigators should be aware of the current optimal methods for handling MPD such as mixed effect models and avoid using simple methods such as LOCF or other single imputation methods that might provoke more biased results[27]. The use of more sophisticated methods is likely to require help from statisticians. Investigators should provide a justification for the sensitivity analyses they choose, and discuss the implication of sensitivity analyses of MPD regarding risk of bias.

Implications for systematic reviewers

When judging the risk of bias of included trials, systematic reviewers should examine the quality and extent of reporting MPD in CONSORT and text of the trials at an outcome of interest level. Furthermore, they should examine the sensitivity analysis regarding MPD conducted in individual trials; this may provide a sense of the extent of risk of bias related to MPD across studies. These results may influence the application of across-trial methods to estimate the impact of MPD on risk of bias across the body of evidence[29,30].

Implications for future research

A checklist addressing the reporting of analysis regarding MPD in RCTs may be useful for both evaluating and optimizing analytic strategies in studies of continuous outcomes. Further investigation might focus on optimal approaches to conducting sensitivity analysis regarding MPD.

Acknowledgement: We thank Lawrence Mbuagbaw for the assistance with data analyses. We also thank Rachel Couban for her assistance with reference uploading with DistillerSR.

Contributors: YZ, EAA, LT, and GHG were responsible for study conception and design. YZ, IDF, LC, SAK, FABA, AL, SC, LCL, TD, WW, JJR, SZ, AA, MW, XJ, GM, YZ, ACL, PA, YF and LAK screened and acquired the data. LM and YZ analyzed the data. YZ drafted the manuscript and is guarantor. All authors critically revised the manuscript and approved the final version.

Funding: This study had no funding support.

Competing interests: None of the authors have either financial or non-financial conflict of interests.

Figure legends:

Figure 1: PRISMA flow diagram

LOST-IT-II 2015 **Flow Diagram**

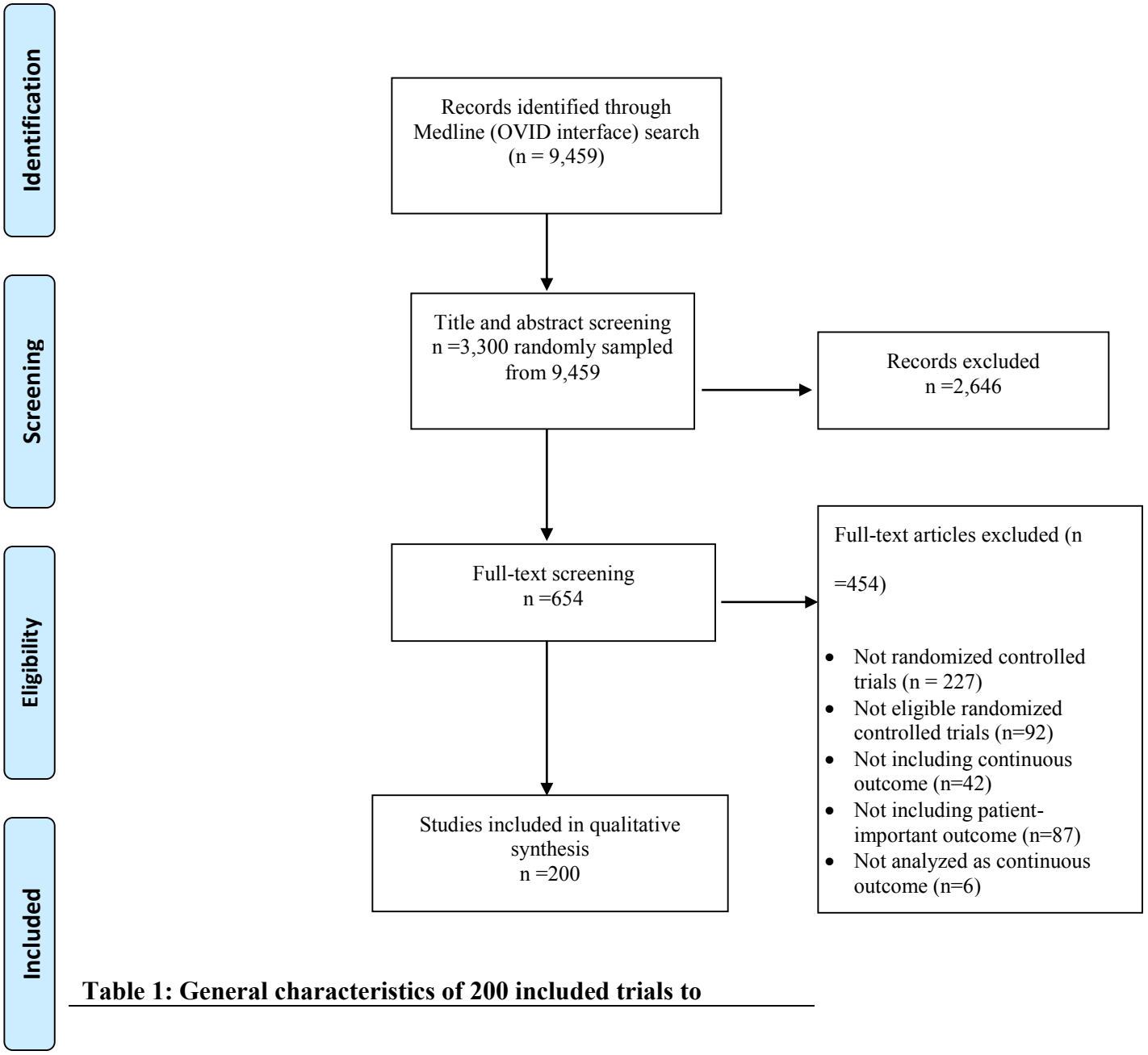

Table 1: General characteristics of 200 included trials to

determine effect of missing participant data on outcomes	
Variable	**n (%)**
Outcome classification	
Efficacy	191 (95.5)
Safety	9 (4.5)
Types of chosen outcome	
Length of stay (in hospital, ICU)	25 (12.5)
Symptoms	84 (42.0)
Quality of life	44 (22.0)
Functional status	33 (16.5)
Disease severity	8 (4.0)
Length of drug use	6 (3.0)
Intervention	
Pharmacological	86 (43.0)
Surgical	24 (12.0)
Invasive non-surgical procedure	14 (7.0)
Rehabilitation	24 (12.0)
Behavioral intervention	24 (12.0)
Diagnostic test	1 (0.5)
Complementary and alternative medicine	3 (1.5)
Other	24 (12.0)
Control	
Standard care	47 (23.5)
Placebo/sham	61 (30.5)
Pharmacological	31 (15.5)
Surgical	16 (8.0)
Invasive non-surgical procedure	7 (3.5)
Rehabilitation	12 (6.0)
Behavioural intervention	10 (5.0)
Diagnostic test	1 (0.5)
Number of centers	
Single center	102 (51.0)
Multi-centers	98 (49.0)
Journal types	
Top 5 journals	37 (18.5)
Non-top 5 journals	163 (81.5)
Arms	
2 arms	154 (77.0)
More than 2 arms	46 (33.0)
Funding*	
Private for profit (ONLY provide drugs)	35 (17.5)
Private for profit (Provide things OTHER than drugs)	38 (19.0)
Private not for profit	72 (36.0)

Governmental	78 (39.0)
Not funded	13 (6.5)
Not reported	25 (12.5)

*Adds up to more than 200 because some trials have more than one source of funding.

Table 2: Methodological characteristics of 200 included trials to determine effect of missing participant data on outcomes

Variable	n (%)
Allocation concealment^	
Adequate	139 (69.5)
Inadequate	61 (30.5)
Blinding*:	
Patients	99 (49.5)
Providers	80 (40.0)
Data collectors	106 (53.0)
Outcome adjudicators	105 (52.5)
Data analysts	21 (10.5)
No early stopping for benefit	**198 (99.0)**
Primary analysis authors described	
Analysis described as Intention to Treat	5 (2.5)
Analysis described as modified Intention to Treat	94 (47.0)
Analysed participants for whom outcome data were available in group to which they were randomised	11 (5.5)
No explicit statement	63 (31.5)
Per protocol analysis	27 (13.5)

^Allocation concealment refers to judgement of "definitely concealed" or " probably concealed"

*Blinding refers to judgment of "definitely blinded" or "probably blinded."

Table 3: Reporting of information regarding missing participant data

in included trials to determine effect of missing participant data on outcomes

Variable	n (%)
Among all included studies (n=200)	
Explicit statement about missing participant data occurred in the main text or CONSORT flow diagram	
Yes, stated MPD occurred	163 (81.5)
Yes, stated MPD did not occur	24 (12.0)
No explicit statement	13 (6.5)
Among studies reported MPD occurred (n=163)	
Assessment of baseline characteristics	
Yes, MPD group vs. non MPD group	12 (7.4)
Yes, MPD in 1st arm vs. MPD in 2nd arm	2 (1.2)
Yes, They did both that mentioned above	1 (0.61)
No	148 (90.8)
Reporting of MPD	
Separately reported for two arms	114 (69.9)
Reported overall only	5 (3.1)
Reported both per arm and overall	44 (27.0)

Table 4: Categories of trial participants trials investigators considered as having MPD

Variable	n (%)
Number of studies stating the different types of MPD (n=176)*	
Ineligible participants/mistakenly randomized	15 (8.5)
Did not receive any intervention	38 (21.5)
Withdrew consent	81 (46.0)
Dead	35 (19.8)
Experienced adverse events	41 (23.3)
Non-compliant /non-adherent	30 (17.0)
Discontinued prematurely	29 (16.5)
Excluded as part of center exclusion	3 (1.7)
Outcome not assessable	14 (7.9)

*The percentages may add up to more than 100 because some trials have more than one category.

Table 5: Planned analytic approach in 200 included trials for MPD on any continuous outcome

Variable	n (%)
Primary analysis	
Yes, planned in methods	58 (29.0)
No, did not plan in the methods	142 (71)
What primary analysis was planned (n=58)	
Complete case analysis	9 (4.5)
All available data analysis	3 (1.5)
Mean imputation	3 (1.5)
Last Observation Carrying Forward (LOCF)	11 (5.5)
Regression for MPD	1 (0.5)
Multiple imputation (MI)	9 (4.5)
Maximum Likelihood (ML)	3 (1.5)
Mixed effect model for missing data	11 (5.5)
Other	6 (3.0)
Not reported	2 (0.0)
Sensitivity analysis	
Yes, planned in methods	21 (10.5)
No, did not plan in the methods	179 (89.5)
What sensitivity analysis was planned (n=21)	
Complete case analysis	3 (1.5)
Mean imputation	1 (0.5)
Last Observation Caring Forward (LOCF)	2 (1.0)
Regression for MPD	1 (0.5)
Multiple imputation (MI)	8 (4.0)
Mixed effect model for missing data	1 (0.5)
Other	3 (1.5)
Not reported	2 (1.0)

Table 6: Reporting of information regarding analysis of missing participant data in included trials on patient-important outcomes

Variable	n (%)
Among studies reported MPD occurred (n=163)	
Assumed missing mechanism when conduct analysis	
Missing at random	13 (7.9)
Ignorable missing	1 (0.7)
Not stated	149 (91.4)
Primary analysis	
Complete case analysis	74 (45.5)
All available data analysis	35 (21.5)
Mean imputation	2 (1.2)
Last Observation Caring Forward (LOCF)	9 (5.5)
Regression for MPD	1 (0.6)
Multiple imputation (MI)	9 (5.5)
Maximum Likelihood (ML)	3 (1.8)
Mixed effect model for missing data	10 (6.1)
Other	4 (2.5)
Unclear	16 (9.8)
Provide justification for the method used to handle MPD in the primary analysis	9 (5.5)
Whether they conducted sensitivity analysis regarding MPD for chosen outcome	
Yes	16 (9.8)
No	147 (90.2)
Implications of MPD regarding risk of bias discussed	18 (11.0)
Among studies conducted sensitivity analysis regarding MPD (n=16) *	
Complete case analysis	3 (1.8)
Mean imputation	1 (0.6)
Last Observation Carrying Forward (LOCF)	2 (1.2)
Multiple imputation (MI)	4 (2.5)
Combination of more than one method above for MPD	1 (0.6)
Other	2 (1.2)
Not reported	6 (4.5)
Sensitivity analysis changed the statistical significant result	2 (1.2)

*(n=16; it is possible that there are more than one sensitivity analysis)

Table 7: Planned sensitivity analysis and conducted sensitivity

analysis in 163 trials with MPD

Variable	n (%)
Sensitivity analysis	
Neither planned nor conducted	138 (84.7)
Planned and conducted	11 (6.7)
Planned but not conducted	9 (5.5)
Did not plan but conducted	5 (3.1)

Reference:

1. Akl EA, Kahale LA, Agarwal A, et al. Impact of missing participant data for dichotomous outcomes on pooled effect estimates in systematic reviews: a protocol for a methodological study. *Syst Rev* 2014; **3**: 137.
2. McKnight PE MK SS, Figueredo AJ. Missing Data: A Gentle Introduction. . New York, NY: Guilford; 2007.
3. te Riele WW, Boerma D, Wiezer MJ, Borel Rinkes IH, van Ramshorst B. Long-term results of laparoscopic adjustable gastric banding in patients lost to follow-up. *Br J Surg* 2010; **97**(10): 1535-40.
4. Geng EH, Emenyonu N, Bwana MB, Glidden DV, Martin JN. Sampling-based approach to determining outcomes of patients lost to follow-up in antiretroviral therapy scale-up programs in Africa. *JAMA* 2008; **300**(5): 506-7.
5. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ* 2001; **165**(10): 1339-41.
6. Little RJ RD. Statistical analysis with missing data. New York: Wiley; 2002.
7. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med* 2013; **86**(3): 343-58.
8. 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.
9. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials* 2004; **1**(4): 368-76.
10. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999; **319**(7211): 670-4.
11. Baron G, Boutron I, Giraudeau B, Ravaud P. Violation of the intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. *Arthritis Rheum* 2005; **52**(6): 1858-65.
12. 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.
13. LOST to follow-up Information in Trials (LOST-IT): a protocol on the potential impact. *Trials* 2009; **10:40.**(doi): 10.1186/745-6215-10-40.
14. Thabane L, Mbuagbaw L, Zhang S, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol* 2013; **13**: 92.
15. Horvitz-Lennon M, James O'Malley A, Frank RG, Normand S-LT. Improving traditional intention-to-treat analyses: a new approach. *Psychological Medicine* 2005; **35**(7): 961-70.
16. Huang X, Jiao L, Wei L, Quan H, Teoh L, Koch GG. Missing radiographic data handling in randomized clinical trials in rheumatoid arthritis. *J Biopharm Stat* 2013; **23**(6): 1435-52.

17. Li X, Wang WW, Liu GF, Chan IS. Handling missing data in vaccine clinical trials for immunogenicity and safety evaluation. *J Biopharm Stat* 2011; **21**(2): 294-310.
18. Liu M, Wei L, Zhang J. Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. *Pharmaceutical Statistics* 2006; **5**(1): 7-18.
19. Tsonaka R, Verbeke G, Lesaffre E. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics* 2009; **65**(1): 81-7.
20. Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Stat Med* 2001; **20**(24): 3931-46.
21. Allison PD. Missing data. Thousand Oaks,CA: Sage; 2002.
22. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; **7**: 10.
23. Sun X, Briel M, Busse JW, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. *BMJ* 2011; **342**: d1569.
24. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005; **294**(17): 2203-9.
25. Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. *JAMA* 2002; **287**(21): 2805-8.
26. Alshurafa M, Briel M, Akl EA, et al. Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. *PLoS One* 2012; **7**(11): e49163.
27. Zhang Y, Alyass, A, Vanniyasingam T.et al. Handling trial participants with missing outcome data for continuous outcomes in randomized control trials: a systematic survey of the methods literature. 2015.
28. Scharfstein DO, Hogan J, Herman A. On the prevention and analysis of missing data in randomized clinical trials: the state of the art. *J Bone Joint Surg Am* 2012; **94 Suppl 1**: 80-4.
29. Ebrahim S, Johnston BC, Akl EA, et al. Addressing continuous data measured with different instruments for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol* 2014; **67**(5): 560-70.
30. Ebrahim S, Akl EA, Mustafa RA, et al. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. *J Clin Epidemiol* 2013; **66**(9): 1014-21 e1.

Appendix1: Hierarchy of outcomes relative to patient importance

I. Mortality
1. All - cause mortality
2. Disease specific mortality
II. Morbidity
1. Cardiovascular major morbid events
2. Other major morbid events (e.g. loss of vision, seizures, fracture, revascularization)
3. Onset/recurrence/relapse/remission of cancer and other chronic diseases (e.g. COPD exacerbation, new onset of diabetes)
4. Renal failure requiring dialysis
5. Hospitalization, medical and surgical procedures (e.g. placement of a pacemaker, and cardioversion)
6. Infections
7. Dermatological/ rheumatologic disorders
III. Symptoms/Quality of life/Functional status (e.g. failure to become pregnant, successful nursing/breastfeeding, depression)
IV. Surrogate outcomes (e.g. viral load, physical activity, weight loss, cognitive function, recurrent polyps, adherence to medication)

Appendix 2: Search strategy for Medline using OVID interface

1. randomized controlled trial.pt.

2. controlled clinical trial.pt.

3. randomized controlled trial/

4. random allocation/

5. double blind method/

6. single blind method/

7. clinical trial.pt.

8. exp clinical trial/

9. exp Clinical Trials as Topic/ or exp Randomized Controlled Trials as Topic/

10. (clin$ adj25 trial$).mp.

11. ((singl$ or doub$ or trebl$ or tripl$) adj25 (blind$ or mask$)).mp.

12. Placebos/

13. placebo$.mp.

14. random$.mp.

15. research design/

16. or/1-15

17. animals/ not humans/

18. 16 not 17

19. "2014".yr.

Appendix 3: multiple linear regression for percentage of missing data

Covariate	Beta coefficient	P-value
Sample size	0.01 (0.0-0.02)	0.005
Intervention type	0.33 (-3.17-3.8)	0.853
Funding	4.89 (0.40-9.38)	0.033
Journal type	-3.05 (-11.12-5.02)	0.455
Allocation concealment	-6.52 (-13.14-0.11)	0.054

(Goodness of fit: R^2=0.167)

CHAPTER 5:

DISCUSSION AND CONCLUSION

Part 1 of this chapter summarizes the research findings from chapters 2 to 4. Part 2 I highlights the implications for trialists, methodologists, and journal editors and discusses directions for future research.

Part 1: Summary of findings

Findings on adherence of checklist on reporting of simulation studies (chapter2)

From 16,446 citations, we found 60 eligible simulation studies addressing statistical methods to deal with continuous missing participant data (MPD) in RCTs. Although almost all studies reported number of simulations (57, 95%) and all studies reported statistical methods compared, scenarios in which they were investigated, and the criteria to assess their relative merits. These simulation studies, the studies nevertheless suffered from limitations in reporting quality. In particular, they omitted details and clarities regarding the procedures of simulations. No study reported starting seeds and only one study reported the random number generator used. Approximately half reported software to perform analysis (29, 48.3%) and less than half reported software to perform simulations (25, 41.7%). Very few reported justification of number of simulation (3, 5%).

Findings on the performance of the statistical methods (chapter3)

We summarized 60 studies that compared 250 methods of dealing with MPD in primary studies using continuous variables of which 47 addressed ignorable (missing completely

at random (MCAR), or missing at random (MAR)) and 32 addressed non-ignorable (not missing at random (NMAR)) data. There was limited overlap on the methods being compared: mixed model was the most frequently investigated class of method (31, 51.7%). We presented the number of times a method performed best and the number of times it performed the worst and the percentage in which it was the best or worst out of the total times it was compared.

Under the ignorable missing assumption, we found mixed model produced the smallest bias (10, 55.6%) and also proved to be the best most frequently on overall ranking (9, 34.6%). However, it had less impressive performance on other properties (precision, accuracy, Type I error, power, and coverage). Multiple imputation (MI) was the other method that stood out with better perfmorance on bias and overall performance than other methods.

Under the non-ignorable assumption mixed model performed the best most frequently on overall ranking (7, 46.7%), bias (8, 57.1%), and accuracy (4, 66.7%) but was less frequently the best on other properties. Multiple imputation performed similarly to mixed model and ranked second most frequently regarding overall performance (6, 42.9%). Last observation caring forward (LOCF) was most frequently the worst on overall ranking (4, 26.7%), bias (7, 46.7%), type I error (3, 75%), and power (3, 75%).

Reporting of analysis of MPD in RCTs (chapter4)

In a representative sample of 200 RCTs published in core medical journals, we found more than 90% (187) of the trials explicitly reported whether MPD occurred or not. Approximately 80% (163) of the trials reported substantial MPD with a median of 11.4%

and interquartile ranges 2.5 to 22.6%. Among the 163 trials explicitly reporting MPD, trialists did not specify how their primary analysis dealt with MPD in 16 (9.8%) trials. Of 147 trials reporting approaches to deal with MPD, 109 (67%) only analyzed only observed data. Of the remainder, trialists augmented data using a mixed effect model (10, 6.1%), or replaced missing value using LOCF (9, 4.5%), multiple imputation (9, 4.5%), maximum likelihood (3, 1.8%), mean imputation (1, 2.0%), regression (1, 0.6%) or another method (4, 2.5%) to deal with MPD. Of the 163 trials with MPD, 18 (11.1%) discussed the implications of MPD regarding risk of bias, and 16 conducted sensitivity analysis to assess the impact of MPD for their primary analysis. Using logistic regression, we found an association between reporting of funding and reporting of MPD, and between larger sample size and longer follow-up time and larger percentage of MPD.

Part 2: Implications for trialists, methodologists and biostatisticians and journal editors

Implications for trialists

Trialists need to be consistent and explicit in reporting MPD at the outcome level and at the patients level. When trials have multiple follow-up times, specific reporting MPD at each follow-up is ideal. Omitting reporting MPD at the outcome level but only at patients level leaves the extent of MPD ambiguous. Trialists might consider applying strategies to minimize MPD and also plan strategies[1] in the protocol as well as the methods section of the manuscript to deal with MPD since MPD is almost always inevitable.

Prior to choosing a method, observing the baseline characteristics and other covariates as well as the observed outcome would lead to making a more plausible assumption about missing mechanism and therefore guide trialists to choose a optimal method. For example, patients living in remote areas may drop out more frequently than patients who live in urban settings. In such an instance, it is more likely that data are MAR[2]. If after considering possible associations between patient characteristics and missingness, the reason for MPD remains unclear, then the most plausible assumption would be NMAR[2].

Trialists may be wise to choose mixed models to deal with MPD rather than other methods irrespective of whether data is ignorable or non-ignorable. Trialists should avoid using LOCF. It might also be helpful to seek help from statisticians to select a specific approach within the categories of mixed model. When choosing methods for sensitivity analysis, trialist should justify their choice. Lastly, they should discuss the implication of MPD on risk of bias.

Implications for methodologists and biostatisticians

It is necessary for statisticians and methodologists to adhere to criteria for transparent and comprehensive reporting to reinforce the clarity and readability of simulation studies assessing performance of statistical methods to deal with continuous MPD in RCTs. Clear description facilitates readers' understanding and increases the reproducibility of such simulations. Moreover, high quality reporting discloses methodological weakness of the simulations, which ease the critical appraisal process and may lead to ultimate improvement in designs and conduct of future research.

Methodologists and statisticians need to be aware of the current available approaches to handling MPD and use the established terminology regarding the name of the methods. When researchers propose new statistical approaches, they should define the category in which the proposed method belongs. Other than considering using our proposed 14-category classification, development of standard classification system might be helpful to ensure the consistency in the literature. Authors should provide the full name along with a clear definition of the method, as well as acronyms when applicable. Further research might be focused on exploring and comparing more superior approaches within one optimal performed category of method such as mixed model. Furthermore, establishing consensus criteria to assess the optimal performance of methods, as well as the procedures to investigate these criteria, is warranted.

When statisticians recommend mixed model to deal with continuous MPD for trialists, they should consider the empirical evidence on simulation study sharing similar characteristics (missing mechanism, sample size, distribution of the data, etc). The 6 simulation studies[3-8] that assessed the performance of the mixed methods, as well as other simulation studies that compared mixed model with other categories of methods- will provide evidence on the selection of method.

Methodologists need to scrutinize the extent of reporting, analytical approaches reported and conducted regarding MPD, as well as the sensitivity analysis conducted to examine the impact of MPD on RCTs when critically apprise the quality of individual RCTs. Our results showed an association between general poor reporting on funding with poor

reporting of MPD. Future research exploring the impact of MPD on statistical important effect estimates may be interesting.

Implications for journal editors

Journal editors need to be aware of the potential limitation and poor reporting quality of simulation studies that deal with MPD. It might be helpful to endorse and recommend that authors use a checklist to enhance the reporting quality to prevent publishing poorly reported simulation studies.

Final remark

This thesis has consistent findings on the extent of MPD with previous investigations on dichotomous MPD in RCTs[9], and commonly applied strategies for MPD in the analysis in RCTs[10]. It created new knowledge synthesis on the reporting quality of simulation investigating methods to deal with continuous MPD for individual RCTs, optimal and inferior statistical methods for handling continuous MPD in RCTs that has systematically tested in simulation studies, and the reporting, analytical approached planned and conducted in RCTs.

Reference:

1. Scharfstein DO, Hogan J, Herman A. On the prevention and analysis of missing data in randomized clinical trials: the state of the art. *J Bone Joint Surg Am* 2012; 94 Suppl 1: 80-4.
2. Little RJ, Rubin DB. Statistical analysis with missing data. New York: Wiley; 2002.
3. Groenwold RH, Donders AR, Roes KC, Harrell FE, Jr., Moons KG. Dealing with missing outcome data in randomized trials and observational studies. Am J Epidemiol. 2012;175(3):210-217.
4. Guo W, Ratcliffe SJ, Have TTT. A Random Pattern-Mixture Model for Longitudinal Data With Dropouts. Journal of the American Statistical Association. 2004;99(468):929-937.
5. Hogan JW, Lin X, Herman B. Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. Biometrics. 2004;60(4):854-864.
6. Li J, Schluchter MD. Conditional mixed models adjusting for non-ignorable drop-out with administrative censoring in longitudinal studies. Stat Med. 2004;23(22):3489-3503.
7. Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Impact of missing data due to selective dropouts in cohort studies and clinical trials. Epidemiology. 2002;13(3):347-355.
8. Yun SC, Lee Y, Kenward MG. Using Hierarchical Likelihood for Missing Data Problems. Biometrika. 2007;94(4):905-919.
9. 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.
10. Alshurafa M, Briel M, Akl EA, et al. Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. *PLoS One* 2012; **7**(11): e49163.