Current and future initiatives in missing data

Katherine Lee, Els Goetghebeur, Rheanna Mainzer and James Carpenter

on behalf of TG1: Missing Data

Aims of TG1

1. Describe the principles for the analysis of partially observed observational data

2. Illustrate potential methods for handling missing data and their application

3. Provide general guidance on how best to handle of missing data across a range of settings

Current Initiatives

Practical guide for sensitivity analyses

- In some scenarios, it may not be possible to estimate a parameter of interest consistently using the observed data alone, i.e. it is not "recoverable"
  - For example, if missingness in a variable depends on the variable itself
- Requires external information about the missing values
  - Need to consider sensitivity analysis regarding the assumptions about the external information
- Important step that is often overlooked and poorly reported (Mainzer et al, 2024)

➢Developing a practical guide on the planning, conduct and reporting of sensitivity analyses
  ➢Raise awareness
  ➢Make more accessible
  ➢Encourage update

Practical guide for sensitivity analyses

Follows the principles
of our TARMOS
framework…
(188 citations since 2021)

1. Plan the analysis
   a) What is the analysis model if no missing data?
   b) How are missing data going to be handled?
- Is a complete records analysis likely to be valid?
- Is MI likely to offer benefits over a complete records analysis?
- Is a sensitivity analysis required?

2. Conduct the analysis
   a) Examine the data – consistent with analysis plan?
   b) Conduct the analysis as per the plan – justifying any amendments

3. Report the analysis
   a) Describe missing data
   b) Describe and justify how missing data were handled
   c) Report all analyses

Practical guide for sensitivity analyses

Follows the principles
of our TARMOS
framework...

(188 citations since 2021)

1. Plan the analysis
   a) What is the analysis model if no missing data?
   b) How are missing data going to be handled?
   - Is a complete records analysis likely to be valid?
   - Is MI likely to offer benefits over a complete records analysis?
   - Is a sensitivity analysis required?

2. Conduct the analysis
   a) Examine the data – consistent with analysis plan?
   b) Conduct the analysis as per the plan – justifying any amendments

3. Report the analysis
   a) Describe missing data
   b) Describe and justify how missing data were handled
   c) Report all analyses

Practical guide for sensitivity analyses

1. Planning the analysis
   1. Which deviations from the assumptions made for the primary analysis need to be considered? i.e. what sensitivity analyses are required?
   2. Which method of estimation?
   3. How to choose the missing values/sensitivity parameter
   4. Writing the analysis plan

2. Conducting the analysis – follow the analysis plan

3. Reporting the sensitivity analysis
   1. Describe the missing data
   2. Presenting the results (graphically)
   3. Interpreting the results – how results differ to the primary results, which is the most believable

Practical guide for sensitivity analyses

Features

➢Accessible for level 1 researcher

➢Provides an overview of analytic approaches

➢Includes a worked case study

  ➢Code for different approaches

  ➢Example text for reporting

➢Checklist for reporting

Case study: Longitudinal Study of Australian Children (LSAC)
1. Proportion overweight age 4-5y
2. Mean Quality of Life at 10-11y
3. Relationship between being overweight and poor quality of life age 10-11y (HRQoL)

Other past initiatives

- A comparison of three popular methods for handling missing data
  - complete case analysis
  - Inverse probability weighting
  - multiple imputation

- Level 1 guidance
  - Describes the 3 approaches
  - Discussion of when each approach is and is not preferable

Future Initiatives

1. Review of journal guidelines

- Ongoing evidence that missing data is often still not handled appropriately despite the guidance available (e.g. Mainzer et al)

- Journals play an important role in increasing quality of papers

- Hardwicke et al (2022) conducted a review of statistical guidelines provided by journals (presented by Willi at ISCB 2023)

- Plan to supplement this with a detailed review of journal guidelines for handling missing data
  - Will start by focusing on the same top ranked medical journals

- Hope to highlighting what is lacking in these guidelines

- Plan to use this develop a template for author guidelines

2. Overview of methods for missing data

- Already published "A comparison of three popular methods for handling missing data: complete case analysis, weighting and multiple imputation" (Little et al, 2022)
- Comprehensive summary of available methods aimed at level 1 researchers
  - Multiple imputation
  - Inverse probability weighting
  - Full Bayesian
  - EM algorithm
  - Maximum likelihood
  - Doubly robust methods

- Discussion of the assumptions needed for each and their plausibility
- Pros and cons of the approaches
- Where one approach may be preferable over another
- Example code for each using a single case study

3. Handling missing data in the context of informative drop out and non-positivity

SISAQOL - Setting International Standards in Analysing patient-reported outcomes and Quality Of Life

- International multidisciplinary consortium, co-led by the European Organization for Research and Treatment of Cancer (EORTC) and Boehringer Ingelheim (BI).
- Convened to generate recommendations to standardize the use, analysis, and interpretation of patient reported outcome (PRO) data in cancer clinical trials.

The data:

o Single arm trial in advanced cancer

o Global Quality of Life (QoL) measured every 3 weeks

o Aim: summarise QoL over time while alive
(other intercurrent events [ICEs] – treatment policy)

SISAQOL | IMI

Setting International Standards in Analysing Patient-Reported
Outcomes and Quality of Life Endpoints

3. Handling missing data in the context of informative drop out and non-positivity

Issues

1. Treatment discontinuation often occurs due to deterioration -> informative drop-out (data missing not at random)

2. Very little/no data post treatment discontinuation -> non-positivity

3. Typical analysis fits a linear mixed model for QoL over time
   - Assumes data are missing at random (conditional on variables in the analysis model)
   - Trends extrapolated post death
   - Time of death and intercurrent events (treatment discontinuation) ignored in implicit imputation of missing values

3. Handling missing data in the context of informative drop out and non-positivity

ISCB 2023: Doranne Thomassen "Imputation of longitudinal patient-reported outcomes in the presence of death and other intercurrent events"

- o Used multiple imputation
- o Imputed until death/censoring (no imputation post death)
- o Incorporated relationship (downward trend) between QoL and ICE/death
- o Time variables modelled with splines
- o Allowed a different relation between QoL and time of ICE/death vs time of censoring (time x censoring interaction)
- o Case study
- o Did not consider the informative missingness

IMPUTATION OF LONGITUDINAL
PATIENT-REPORTED OUTCOMES
IN THE PRESENCE OF DEATH
AND OTHER INTERCURRENT
EVENTS

Doranne Thomassen (LUMC), Satrajit Roychoudhury (Pfizer Inc.), Cecilie
Delphin Amdal (UHOslo), Limin Liu (UGent), Jammbe Z. Musoro (EORTC), Willi
Sauerbrei (UFreiburg), Saskia le Cessie (LUMC), Els Goetghebeur (UGent);
on behalf of SISAQOL-IMI Work Package 3
Contact: d.thomassen@lumc.nl

3. Handling missing data in the context of informative drop out and non-positivity

Possible papers*:

1. Discussion of the issues and methods for handling them (level 1 researchers)
    a. Informative censoring
    b. Non-positivity

    Descriptive case study (based on Doranne's work)

2. Comparison of methods (level 2 researchers)
    a. Different imputation approaches
    b. Different methods of estimating survival
    c. Methods for handling of non-positivity and informative missingness (sensitivity analyses)

    Simulation study (and a case study)

* In collaboration with TG7: causal inference

4. Other initiatives in the pipeline

- Joint project with TG4 (measurement error) on simultaneous handling of measurement error and missing data

- Joint project with Dr Ellie Curnow (Bristol University) to develop `midoc`, an expert system to guide analysts through analysing a dataset with missing observations (prototype presented at ISCB in 2023)

Website: https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#stratos

STRATOS TG1 Members

- Kate Lee, Murdoch Children's Research Institute & University of Melbourne, Australia
- James Carpenter, London School of Hygiene and Tropical Medicine, London, UK
- Melanie Bell, University of Arizona, USA
- Rosie Cornish, University of Bristol, Bristol, UK
- Els Goetghebeur, Ghent University, Belgium
- Joseph Hogan, Brown University, USA
- Rod Little, University of Michigan, USA
- Rheanna Mainzer, Murdoch Children's Research Institute, Melbourne, Australia
- Kate Tilling, University of Bristol, Bristol, UK

Reference

- Lee, K. J, Tilling, K, Cornish, R. P, Little, R. J. A, Bell, M. L, Goetghebeur, E., Hogan J.W. and Carpenter, J. R., on behalf of the STRATOS initiative (2020). Framework for the treatment and reporting of missing data in observational studies: the TARMOS framework. JCE 2021. 134: P79-88. 2021 Jun;134:79-88. doi: 10.1016/j.jclinepi.2021.01.008.

- Litte RJ, Carpenter JR and Lee KJ on behalf of the STRATOS initiative. A comparison of three popular methods for handling missing data: complete case analysis, weighting and multiple imputation. Sociological Methods & Research, 2022. https://doi.org/10.1177/004912412211138.

- Mainzer RM, Lee KJ, Tilling KM, Cornish RPm Moreno-Betancur M, Goetghebeur E and Carpenter J, Little RJA, Bell ML, Hogan JW on behalf of the STRATOS initiative JR. Sensitivity analyses for missing data: A practical guide for planning, conducting and reporting. Nearly submitted!