rbmi: A R package for standard and reference-based multiple imputation methods

Craig Gower-Page* Alessandro Noci† Marcel Wolbers†

January 12, 2022

Summary

Many randomized controlled clinical trials compare a continuous outcome variable which is assessed longitudinally at scheduled follow-up visits between the subjects assigned to the intervention treatment and those assigned to the control group. Missing outcome measurements may occur because subjects miss an assessment or drop out from the trial altogether. Moreover, intercurrent events (ICEs) such as discontinuations of the assigned treatment or initiations of rescue medications may affect the interpretation or the existence of the outcome measurements associated with the clinical question of interest. The ICH E9(R1) addendum on estimands, a regulatory document published by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, presents a structured framework to link trial objectives to a precise description of the targeted treatment effect in the presence of ICEs and missing data [ICH E9 working group, 2019].

The R package `rbmi` was created to support trial analyses which are aligned with the estimands framework. Missing data is handled using multiple imputation (MI) assuming multivariate normally distributed data. The package supports both standard imputation under a missing-at-random assumption and reference-based imputation methods. Reference-based methods impute missing data in the intervention treatment group based on observed data from the control group [Carpenter et al., 2013]. δ-based imputation methods which add an offset term, δ, to the imputed values prior to the analysis in order to assess the impact of unobserved outcomes being worse or better than those observed are also supported. Such methods are frequently used for sensitivity or "tipping point" analyses [Cro et al., 2020].

Statement of need

`rbmi` is a flexible `R` package designed to support the analysis of randomized clinical trials with continuous longitudinal endpoints. Both conventional MI methods based on Bayesian posterior draws and novel methods based on maximum likelihood estimation and re-sampling are implemented [von Hippel and Bartlett, 2021, Wolbers et al. [2021]]. `rbmi` was designed for statisticians from both academic clinical research units and pharmaceutical industry. To our knowledge, a comprehensive and fully validated `R` implementation of such approaches is still lacking. An established software implementation of reference-based imputation in SAS are the so-called "five macros" [Roger, 2021]. An alternative `R` implementation which is currently under development is the R package `RefBasedMI`[McGrath and White, 2021].

*Data & Statistical Sciences, Pharma Development, Roche, Welwyn Garden City, UK
†Data & Statistical Sciences, Pharma Development, Roche, Basel, Switzerland

Implementation

All approaches implemented in `rbmi` follow a common workflow based on 4 core functions which are called sequentially:

- `draws()` - fits the imputation models and stores their parameters
- `impute()` - creates multiple imputed datasets
- `analyse()` - analyses each of the multiple imputed datasets
- `pool()` - combines the analysis results across imputed datasets into a single statistic

This modular design creates a user-friendly and extensible environment that allow the user to have a direct control on all the phases of the estimation process. In addition, a variety of helper functions have been implemented to further support the user.

The `draws()` function has 3 input arguments:

- `data`: The primary longitudinal `data.frame` containing the outcome variable and all covariates. The inclusion of time-varying covariates is also possible.
- `data_ice`: A `data.frame` which specifies the first visit affected by an ICE and the imputation strategy for handling missing outcome data after the ICE.
- `method`: The selected statistical approaches which may be defined by creating a `method` object using the functions:
  - `method_bayes()` for MI based on Bayesian posterior parameter draws from MCMC sampling and inference based on Rubin's rules [Carpenter et al., 2013].
  - `method_approxbayes()`: as for `method_bayes()` except that approximate Bayesian posterior draws are obtained via bootstrapping and maximum likelihood estimation (Little and Rubin [2002, Section 10.2.3, part 6]).
  - `method_condmean()` for conditional mean imputation based on maximum likelihood estimation. Inference is based on re-sampling techniques (bootstrap or jackknife) as described in Wolbers et al. [2021].
  - `method_bmlmi()` for bootstrapped maximum likelihood MI as described in von Hippel and Bartlett [2021].

In addition to detailed help files for all functions, the package contains three vignette: a `quickstart` vignette which describes the basic functionality, an `advanced` vignette which describes some of the advanced features, and a `stat_specs` vignette which describes the statistical methodology in detail.

Availability and validation

`rbmi` is developed open source on https://github.com/insightsengineering/rbmi and major releases will also be uploaded to CRAN. Source code can only be promoted to the master branch after review by a second programmer. Unit tests which define and document the expected input and output of each function have been implemented to ensure that the package performs as expected. To date, `rbmi` has been used in two simulation studies reported in Wolbers et al. [2021] and Noci et al. [2021].

Acknowledgements

The authors thank Jonathan Bartlett from the University of Bath and Paul Delmar and Daniel Sabanés Bové from Roche for many helpful discussions on the statistical methodology and the software implementation.

References

James R Carpenter, James H Roger, and Michael G Kenward. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics*, 23(6):1352–1371, 2013.

Suzie Cro, Tim P Morris, Michael G Kenward, and James R Carpenter. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. *Statistics in Medicine*, 39(21):2815–2842, 2020.

ICH E9 working group. ICH E9 (R1): Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials., 2019. URL https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data, Second Edition*. John Wiley & Sons, 2002.

Kevin McGrath and Ian White. Refbasedmi: Reference-based imputation for longitudinal clinical trials with protocol deviation, 2021. URL https://github.com/UCL/RefbasedMI. version 0.0.18.

Alessandro Noci, Marcel Wolbers, Markus Abt, Corine Baayen, Hans Ulrich Burger, Man Jin, and Weining Zhao Robieson. A comparison of estimand and estimation strategies for clinical trials in early parkinson's disease, 2021. URL https://arxiv.org/abs/2112.03700.

James Roger. Reference-based mi via multivariate normal rm (the "five macros" and miwithd), 2021. URL https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-missing-data.

Paul T von Hippel and Jonathan W Bartlett. Maximum likelihood multiple imputation: Faster imputations and consistent standard errors without posterior draws. *Statistical Science*, 36(3):400–420, 2021.

Marcel Wolbers, Alessandro Noci, Paul Delmar, Craig Gower-Page, Sean Yiu, and Jonathan W Bartlett. Reference-based imputation methods based on conditional mean imputation, 2021. URL https://arxiv.org/abs/2109.11162.