The Journal of Open Source Software

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

Craig Gower-Page[1], **Alessandro Noci**[2], **and Marcel Wolbers**[2]

1 Data and Statistical Sciences, Pharma Development, Roche, Welwyn Garden City, UK **2** Data and Statistical Sciences, Pharma Development, Roche, Basel, Switzerland

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

Many randomized controlled clinical trials compare a continuous outcome variable that is assessed longitudinally at scheduled follow-up visits between subjects assigned to a intervention treatment group and those assigned to a 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 analyses of trial endpoints which are aligned with the defined estimand. 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 a reference group which is typically defined as the control group of the trial (Carpenter, Roger, and Kenward 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 resampling (as described in von Hippel and Bartlett (2021) and Wolbers et al. (2022)) are implemented. `rbmi` was designed for statisticians from both academic clinical research units and pharmaceutical industry. To our knowledge, a comprehensive and fully unit-tested `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).

`rbmi` workflow

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 allows the user to have direct control on all 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. Imputation of missing data prior to any ICE is always under a missing-at-random assumption.
- `method`: The selected statistical approach which is defined by creating a `method` object by using one of:
  - `method_bayes()` for MI based on Bayesian posterior parameter draws from MCMC sampling and inference based on Rubin's rules (Carpenter, Roger, and Kenward 2013).
  - `method_approxbayes()`: as for `method_bayes()` except that approximate Bayesian posterior draws are obtained via bootstrapping and maximum likelihood estimation (Little and Rubin (2002Section 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. (2022).
  - `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 vignettes: 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.

Availability and testing

`rbmi` is developed open source on https://github.com/insightsengineering/rbmi and major releases will also be uploaded to CRAN. All production code is required to have been reviewed by an independent programmer as well as pass a suite of automated unit tests which both define and document the expected input and output of each function. Additionally comparisons are made to similar software (namely the so-called "five macros" (Roger 2021) SAS implementation) to ensure consistency of results as well as to simulated datasets with known values. To date, `rbmi` has been used in two simulation studies reported in Wolbers et al. (2022) and Noci et al. (2021).

The Journal of Open Source Software

Author contributions and acknowledgements

Craig Gower-Page and Alessandro Noci are the primary developers of the `rbmi` package. Marcel Wolbers initiated the project (jointly with Paul Delmar), specified the statistical methods and contributed to the documentation and vignettes.

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

Carpenter, James R, James H Roger, and Michael G Kenward. 2013. "Analysis of Longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation." *Journal of Biopharmaceutical Statistics* 23 (6): 1352–71. https://doi.org/10.1080/10543406.2013.834911.

Cro, Suzie, Tim P Morris, Michael G Kenward, and James R Carpenter. 2020. "Sensitivity Analysis for Clinical Trials with Missing Continuous Outcome Data Using Controlled Multiple Imputation: A Practical Guide." *Statistics in Medicine* 39 (21): 2815–42. https://doi.org/10.1002/sim.8569.

ICH E9 working group. 2019. "ICH E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials." International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. 2019. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

Little, Roderick JA, and Donald B Rubin. 2002. *Statistical Analysis with Missing Data, Second Edition*. John Wiley & Sons. https://doi.org/10.1002/9781119013563.

McGrath, Kevin, and Ian White. 2021. "RefBasedMI: Reference-Based Imputation for Longitudinal Clinical Trials with Protocol Deviation." https://github.com/UCL/RefbasedMI.

Noci, Alessandro, Marcel Wolbers, Markus Abt, Corine Baayen, Hans Ulrich Burger, Man Jin, and Weining Zhao Robieson. 2021. "A Comparison of Estimand and Estimation Strategies for Clinical Trials in Early Parkinson's Disease." https://arxiv.org/abs/2112.03700.

Roger, James. 2021. "Reference-Based MI via Multivariate Normal RM (the 'Five Macros' and MIWithD)." https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-missing-data.

von Hippel, Paul T, and Jonathan W Bartlett. 2021. "Maximum Likelihood Multiple Imputation: Faster Imputations and Consistent Standard Errors Without Posterior Draws." *Statistical Science* 36 (3): 400–420. https://doi.org/10.1214/20-STS793.

Wolbers, Marcel, Alessandro Noci, Paul Delmar, Craig Gower-Page, Sean Yiu, and Jonathan W Bartlett. 2022. "Standard and Reference-Based Conditional Mean Imputation." https://arxiv.org/abs/2109.11162.