

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

Craig Gower-Page¹, Alessandro Noci², and Marcel Wolbers²

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

DOI: [10.xxxxxx/draft](https://doi.org/10.1093/joss/0000000000000000)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: ↗

Submitted: 02 March 2022

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

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 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 (as described in von Hippel & Bartlett (2021) and Wolbers et al. (2021)) 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 & 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 et al., 2013).
 - `method_approxbayes()`: as for `method_bayes()` except that approximate Bayesian posterior draws are obtained via bootstrapping and maximum likelihood estimation (Little & Rubin (2002, sec. 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 & 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. (2021) and Noci et al. (2021).

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.

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

86 References

- 87 Carpenter, J. R., Roger, J. H., & Kenward, M. G. (2013). Analysis of longitudinal trials with
88 protocol deviation: A framework for relevant, accessible assumptions, and inference via
89 multiple imputation. *Journal of Biopharmaceutical Statistics*, 23(6), 1352–1371.
- 90 Cro, S., Morris, T. P., Kenward, M. G., & Carpenter, J. R. (2020). Sensitivity analysis for
91 clinical trials with missing continuous outcome data using controlled multiple imputation:
92 A practical guide. *Statistics in Medicine*, 39(21), 2815–2842.
- 93 ICH E9 working group. (2019). *ICH E9(R1) Statistical Principles for Clinical Trials: Ad-*
94 *dendum: Estimands and Sensitivity Analysis in Clinical Trials*. International Council for
95 Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. [https://](https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf)
96 database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf
- 97 Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*, second edition.
98 John Wiley & Sons.
- 99 McGrath, K., & White, I. (2021). *RefBasedMI: Reference-based imputation for longitudinal*
100 *clinical trials with protocol deviation*. <https://github.com/UCL/RefbasedMI>
- 101 Noci, A., Wolbers, M., Abt, M., Baayen, C., Burger, H. U., Jin, M., & Robieson, W. Z. (2021).
102 *A comparison of estimand and estimation strategies for clinical trials in early parkinson's*
103 *disease*. <https://arxiv.org/abs/2112.03700>
- 104 Roger, J. (2021). *Reference-based MI via multivariate normal RM (the “five macros” and*
105 *MIWithD)*. [https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#](https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-missing-data)
106 [dia-missing-data](https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-missing-data)
- 107 von Hippel, P. T., & Bartlett, J. W. (2021). Maximum likelihood multiple imputation: Faster
108 imputations and consistent standard errors without posterior draws. *Statistical Science*,
109 36(3), 400–420.
- 110 Wolbers, M., Noci, A., Delmar, P., Gower-Page, C., Yiu, S., & Bartlett, J. W. (2021).
111 *Reference-based imputation methods based on conditional mean imputation*. [https://](https://arxiv.org/abs/2109.11162)
112 arxiv.org/abs/2109.11162