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

Alessandro Noci*, Craig Gower-Page†, Marcel Wolbers‡

December 22, 2021

1 Summary

In clinical trials the estimand framework is a systematic approach to define in detail what needs to be estimated to address the scientific question of interest posed by the trial objective. This framework has been proposed in the ICH E9(R1) addendum on estimands and sensitivity analyses to ensure alignment among clinical trial objectives, trial execution/conduct, statistical analyses, and interpretation of results [ICH E9 working group, 2019].

Once the estimand has been defined, the evaluation of the efficacy of the treatment under investigation requires the definition of an estimator which is well-aligned with the estimand. For more than a decade a mixed effects model repeated measures (MMRM) approach has been the de facto standard for the primary analysis when the primary outcome is a continuous longitudinal variable. However, it is increasingly recognized that alternative statistical methods to MMRM may be required in order to properly align the analysis method with the targeted estimand. A recent overview of methods to align the estimator with the estimand is Mallinckrodt et al. [2020]. A short introduction on estimation methods for studies with longitudinal endpoints can also be found in Wolbers et al. [2021]. One prominent statistical method for this purpose is multiple imputation (MI), which is the target of the `rbmi` package.

Methods based on MI are an attractive and flexible alternative to MMRM. They allow for the imputation of missing data under both missing-at-random (MAR) and missing-not-at-random (MNAR) assumptions. For example, they allow imputation of missing data after a treatment discontinuation in the active arm based on data from the control arm (so-called reference-based imputation). Different reference-based assumptions can be adopted depending on the disease area and the anticipated mechanism of action of the intervention. For a general description and review of reference-based imputation methods, we refer to Carpenter et al. [2013], Cro et al. [2020], and Wolbers et al. [2021].

The ICH E9(R1) addendum also stresses the importance of performing proper sensitivity analyses to assess how robust the results given by the primary analysis are [ICH E9 working group, 2019]. Sensitivity analyses should also be planned to evaluate the impact of the modeling assumptions about missing data. For example, one could penalize the mean outcome of subjects with unobserved data by a pre-specified fixed amount. This kind of approach is called δ-adjustment [Cro et al., 2020], where δ refers to the amount of the penalization. This method can be used to perform a "tipping-point analysis", which consists in progressively increasing the amount of δ adjustment from 0 until the conclusions from the primary analysis are overturned. Sensitivity analyses should also be performed for MI-based estimators and `rbmi` supports this.

---

*Affiliation: F.Hoffmann La-Roche
†Affiliation: F.Hoffmann La-Roche
‡Affiliation: F.Hoffmann La-Roche

2 Statement of need

`rbmi` is a flexible R package designed to support the analysis of randomized clinical trials with continuous longitudinal endpoints. It allows for imputation under missing-at-random and under reference-based methods. Both conventional MI methods based on Bayesian posterior draws and novel methods based on conditional mean imputation and re-sampling are supported. `rbmi` also includes the possibility of performing sensitivity analyses based on fixed δ-adjustments. `rbmi` was designed for statisticians from both academic research units and pharmaceutical industry who deal with late-phase randomized clinical trials. `rbmi` could be used for regulatory purposes or in simulation studies. To our knowledge, a comprehensive, reliable and fully validated implementation of such approaches that also serves for regulatory purposes is still missing. An established software implementation of reference-based imputation in SAS are the so-called "five macros" by James Roger [Roger, 2021]. An alternative `R` implementation which is also currently under development is the R package `RefBasedMI`[McGrath and White, 2021].

3 Implementation

All the approaches implemented in `rbmi` follow a common workflow based on 4 steps whose corresponding functions should be called sequentially. Additionally, helper functions that simplify the set-up of the input parameters and the interpretation of the outputs have also been implemented. This creates a user-friendly environment that allow the user to have a direct control on all the phases of the estimation process. The 4 core functions are the following:

- `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.

The different approaches can be set by defining a `method` object using the functions:

- `method_bayes()` for conventional Bayesian MI based on MCMC sampling and random imputation. Inference is based on Rubin's rules (Barnard and Rubin [1999], Little and Rubin [2002]). This method is described in Carpenter et al. [2013] and Cro et al. [2020].
- `method_approxbayes()` for approximate Bayesian MI where posterior draws are obtained via bootstrap instead of via MCMC (Little and Rubin [2002], Efron [1994], Honaker and King [2010], von Hippel and Bartlett [2021]).
- `method_condmean()` for maximum likelihood parameter estimation and conditional mean imputation. Inference is based on re-sampling techniques (bootstrap [Efron and Tibshirani, 1994] or jackknife [Efron and Stein, 1981]). This method is described in Wolbers et al. [2021].
- `method_bmlmi()` for bootstrapped maximum likelihood MI as described in von Hippel and Bartlett [2021].

A detailed description and comparison of the implemented methods can be found in Wolbers et al. [2021], as well as in the package vignette describing the statistical specifications of the package.

4 Acknowledgements

TODO

References

John Barnard and Donald B Rubin. Miscellanea. small-sample degrees of freedom with multiple imputation. *Biometrika*, 86(4):948–955, 1999.

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.

Bradley Efron. Missing data, imputation, and the bootstrap. *Journal of the American Statistical Association*, 89(426):463–475, 1994.

Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596, 1981.

Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

James Honaker and Gary King. What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581, 2010.

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.

CH Mallinckrodt, J Bell, G Liu, B Ratitch, M O'Kelly, I Lipkovich, P Singh, L Xu, and G Molenberghs. Aligning estimators with estimands in clinical trials: putting the ich e9 (r1) guidelines into practice. *Therapeutic innovation & regulatory science*, 54(2):353–364, 2020.

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.

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.