Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *doi: 10.18637/jss.v000.i00*

ShinyMICE: An Evaluation Suite for Multiple Imputation

Hanne Oberman
Utrecht University

Abstract

The goal of this Master's thesis was to develop novel methodology and guidelines for evaluating multiple imputation methods, and implement these in an interactive evaluation framework for multiple imputation: **ShinyMICE**.

Keywords: multiple imputation, evaluation methodology, **ShinyMICE**, **mice**, R.

1. Introduction: Multiple Imputation Methodology

[**From thesis proposal:**] At some point, any scientist conducting statistical analyses will run into a missing data problem (Allison 2001). Missingness is problematic because statistical inference cannot be performed on incomplete data, and ad hoc solutions can yield wildly invalid results (Van Buuren 2018). To circumvent the ubiquitous problem of missing information, Rubin (1987) proposed the framework of multiple imputation (MI). MI is an iterative algorithmic procedure in which missing data points are 'guessed' (i.e. imputed) several times. The variability between the imputations validly reflects how much uncertainty in the inference is due to missing information–that is, if all statistical assumptions are met Rubin (1987).

[**From thesis proposal:**] With MI, many assumptions are made about the nature of the observed and missing parts of the data and their relation to the 'true' *data generating model* (Van Buuren 2018). Without proper evaluation of the imputations and the underlying assumptions, any drawn inference may erroneously be deemed valid. Such evaluation measures are currently missing or under-developed in MI software, like the world leading R package **mice** (Van Buuren and Groothuis-Oudshoorn 2011). Therefore, I will answer the following question: 'Which measures are vital for evaluating the validity of multiply imputed data?'.

1.1. Features

The aim of this paper is to provide applied researchers an introduction to the interactive evaluation device **ShinyMICE**. The intended audience consists of empirical researchers and statisticians who (want to) use multiple imputation to solve missing data problems. Basic familiarity with multiple imputation methodology is assumed. For an accessible and comprehensive introduction to MI from an applied perspective, see Van Buuren (2018). For the theoretical foundation of MI, see Rubin (1987). All programming code used in this paper is available in the file 'XYZ.R' along with the manuscript, and on Github repository 'XYZ'.

1.2. Notation

The R **Shiny** application introduced in this paper is developed with the aim to integrate it into the **mice** environment. **ShinyMICE** therefore follows notation and conventions of Van Buuren and Groothuis-Oudshoorn (2011). For an overview of the deviations from the 'original' notation by Rubin (1987), see Van Buuren (2018).

Let Y denote an $n \times p$ matrix containing the data values on p variables for all n units in a sample. The collection of observed data values is denoted as Y_{obs}; the missing part of Y is refered to as Y_{mis}. Response indicator R shows whether a data value in Y is missing or observed. The relation between R, Y_{obs}, and Y_{mis} determines the missingness mechanism.

Terminology (MCAR, MAR, MNAR).

Blue points are observed, the red points are imputed.

2. Theoretical Background

- missingness mechanims, ignorability

"The practical importance of the distinction between MCAR, MAR and MNAR is that it clarifies the conditions under which we can accurately estimate the scientifically interesting parameters without the need to know ψ" (Van Buuren 2018, par. 2.2).

- Rubin's rules

- FCS vs. JM?

[**From thesis proposal:**] The validity of the MI solution depends on numerous assumptions that cannot be verified from the observed data alone. So instead of statistical tests for assumptions, evaluation procedures have been developed. For the following assumptions, no reliable procedure has been proposed and/or implemented: 1) *ignorability* of the *missingness mechanism* (Rubin 1987); 2) *congeneality* of the imputation models (Meng 1994); and 3) *compatibility* of the MI modeling procedure (Rubin 1996).

1. [**From thesis proposal:**] A missingness mechanism is said to be ignorable when the probability to be missing does not depend on the missing data itself. Violation of this assumption can gravely affect inferences. Robustness of inferences to varying degrees of violation can be assessed with sensitivity analyses. Some practical guidelines exist (e.g., (Nguyen, Carlin, and Lee 2017)), but current MI software does not facilitate this methodology for empirical researchers.

2. [**From thesis proposal:**] Congeneal imputation models capture all required relations between observed and missing parts of the data. The extent to which this has been

successful can be evaluated by plotting conditional distributions (Abayomi, Gelman, and Levy 2008). Such visualizations are available in MICE, but subsequent statistical tests to quantify the relations with covariates are not provided.

3. **[From thesis proposal:]** The third assumption is met when the MI algorithm converges to a stable distribution. However, conventional measures to diagnose convergence–e.g., Gelman and Rubin's 1992 statistic $\widehat{R}$ –are not applicable on multiply imputed data (Lacerda, Ardington, and Leibbrandt 2007). Therefore, empirical researchers have to rely on visual inspection procedures that are theoretically equivalent to $\widehat{R}$ (White, Royston, and Wood 2011). Visually assessing convergence is not only difficult to the untrained eye, it might also be futile. The convergence properties of MI algorithms lack scientific consensus (Takahashi 2017), and some default MICE techniques might not converge to stable distributions at all (Murray 2018). Moreover, convergence diagnostics for MI methods have not been systematically studied (Van Buuren 2018).

[From thesis proposal:] In short, the existing literature provides both possibilities and limitations to evaluating the validity of multiply imputed data. The goal of this research project is to develop novel methodology and guidelines for evaluating MI methods, and implement these in an interactive evaluation framework for multiple imputation. This framework will aid applied researchers in drawing valid inference from incomplete datasets.

3. Methods

[From thesis proposal:] Initially, the research project will consist of an investigation into algorithmic convergence of MI algorithms. I will replicate Lacerda et al.'s simulation study on $\widehat{R}$ (Lacerda *et al.* 2007), and develop novel guidelines for assessing convergence. Ideally, I will integrate several diagnostics (e.g., $\widehat{R}$, *auto-correlation*, and *simulation error*) into a single summary indicator to flag non-convergence.

[From thesis proposal:] Subsequently, I will use R Shiny (Chang, Cheng, Allaire, Xie, and McPherson 2017) to implement the convergence indicator and existing evaluation measures in **ShinyMICE**, see Figure 1. The application will at least contain methodology for: sensitivity analyses; data visualizations (e.g., scatter-plots, densities, cross-tabulations); and statistical evaluation of relations between variables pre- versus post-imputation (i.e., χ^2 tests or t tests).

[From thesis proposal:] A working beta version of **ShinyMICE** will be considered a sufficient milestone to proceed with writing a technical paper on the methodology and the software. I will submit the paper for publication in *Journal of Statistical Software*. Finally, **ShinyMICE** will be integrated into the existing MICE environment, and a vignette for applied researchers will be written. The R code and documentation of this project will be open source (available on Github). Since the study does not require the use of unpublished empirical data, I expect that the FETC will grant the label exempt.

3.1. Sensitivity Analyses

What is already implemented?

- Current guideine: "Plot densities of both the observed and imputed values of all variables to see whether the imputations are reasonable. Differences in the densities between the observed and imputed values may suggest a problem that needs to be further checked" (Van Buuren and Groothuis-Oudshoorn 2011, p. 43).

What is not yet implemented, but exists?

- Robustness of inferences to varying degrees of violation can be assessed with sensitivity analyses. See (Nguyen *et al.* 2017) for guidelines.
- Forked MICE with MNAR sensitivity analyses

What is not implemented, and NA?

- Nothing within the scope of this thesis.

3.2. Data Visualization

What is already implemented?

- Existing plot functions to evaluate relations between observed and missing parts of the data (**mice** functions `hist()`, `xyplot()`, and `stripplot()`).

What is not yet implemented, but exists?

- Improve plots using **lattice** package functionalities.
- Subsequent statistical tests to quantify the relations with covariates.
- Use evaluation of MI methods in SAS as example

What is not implemented, and NA?

- Nothing within the scope of this thesis.

3.3. Convergence Diagnostic

What is already implemented?

- Traceplots

What is not yet implemented, but exists?

- $\widehat{R}$, but too stringent (new) threshold, and assumption of overdispersed initial values of imputation chains not met.
- Autocorrelation. Schafer (1997, p. 129) wrote on worst linear statistic. We could calculate the autocorrelation of that statistic to know that the algorithm converged elsewhere too. See autocorr function plot in sas of worst linear function.
- Sensitivity analysis: Run algorithm several times and compare results.

What is not implemented, and NA?

- $\hat{R}$ threshold: Replicate simulation study and build a decision rule to solve the problem with

$\hat{R}$.

- Stability of the solution: Possibly use the slope of means over iterations too to see whether there is trending. Or apply PCA on the imputed data and if that (the eigenvalues?) stays the same we know that the means and variances are stable as well, see McKay (?).

- MC error: MC error = SD/sqrt(number of iterations), where SD represents the variation across iterations. The MC error thus represents how much the means differ w.r.t. the iterations. MC error decreases as number of iterations increases. It should not be larger than 5% of the sample standard deviation.

Computational details

The results in this paper were obtained using R 3.6.2 with the **mice** 3.7.0 package. R itself and all packages used are available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/.

Acknowledgments

This paper is written by the sole author (Hanne Oberman, BSc.), with guidance from Master thesis supervisors prof. dr. Stef van Buuren, and dr. Gerko Vink.

References

Abayomi K, Gelman A, Levy M (2008). "Diagnostics for multivariate imputations." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **57**(3), 273–291. ISSN 0035-9254, 1467-9876. doi:10.1111/j.1467-9876.2007.00613.x. URL http://doi.wiley.com/10.1111/j.1467-9876.2007.00613.x.

Allison PD (2001). *Missing data*, volume 136. Sage publications.

Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2017). "Shiny: web application framework for R." URL https://CRAN.R-project.org/package=shiny.

Gelman A, Rubin DB (1992). "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*, **7**(4), 457–472. ISSN 0883-4237, 2168-8745. doi:10.1214/ss/1177011136. URL https://projecteuclid.org/euclid.ss/1177011136.

Lacerda M, Ardington C, Leibbrandt M (2007). "Sequential regression multiple imputation for incomplete multivariate data using Markov chain Monte Carlo."

Meng XL (1994). "Multiple-Imputation Inferences with Uncongenial Sources of Input." *Statistical Science*, **9**(4), 538–558. ISSN 0883-4237, 2168-8745. doi:10.1214/ss/1177010269. URL https://projecteuclid.org/euclid.ss/1177010269.

Murray JS (2018). "Multiple Imputation: A Review of Practical and Theoretical Findings." *Statistical Science*, **33**(2), 142–159. ISSN 0883-4237. doi:10.1214/18-STS644. URL https://projecteuclid.org/euclid.ss/1525313139.

Nguyen CD, Carlin JB, Lee KJ (2017). "Model checking in multiple imputation: an overview and case study." *Emerging Themes in Epidemiology*, **14**(1), 8. ISSN 1742-7622. doi:10.1186/s12982-017-0062-6. URL https://doi.org/10.1186/s12982-017-0062-6.

Rubin DB (1987). *Multiple Imputation for nonresponse in surveys*. Wiley series in probability and mathematical statistics Applied probability and statistics. Wiley, New York, NY. ISBN 978-0-471-08705-2.

Rubin DB (1996). "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association*, **91**(434), 473–489. ISSN 01621459. doi:10.2307/2291635. URL http://www.jstor.org/stable/2291635.

Takahashi M (2017). "Statistical Inference in Missing Data by MCMC and Non-MCMC Multiple Imputation Algorithms: Assessing the Effects of Between-Imputation Iterations." *Data Science Journal*, **16**, 37. ISSN 1683-1470. doi:10.5334/dsj-2017-037. URL http://datascience.codata.org/articles/10.5334/dsj-2017-037/.

Van Buuren S (2018). *Flexible imputation of missing data*. Chapman and Hall/CRC. ISBN 0-429-96035-2.

Van Buuren S, Groothuis-Oudshoorn K (2011). "mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software*, **45**(1), 1–67. ISSN 1548-7660. doi:10.18637/jss.v045.i03. URL https://www.jstatsoft.org/index.php/jss/article/view/v045i03.

White IR, Royston P, Wood AM (2011). "Multiple imputation using chained equations: Issues and guidance for practice." *Statistics in Medicine*, **30**(4), 377–399. ISSN 1097-0258. doi:10.1002/sim.4067.

Affiliation:

Hanne Ida Oberman, BSc.
Methodology and Statistics for the Behavioural, Biomedical and Social Sciences
Department of Methodology and Statistics
Faculty of Social and Behavioral Sciences
Utrecht University
Heidelberglaan 15
3500 Utrecht, The Netherlands
E-mail: h.i.oberman@uu.nl

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:10.18637/jss.v000.i00 *Accepted:* yyyy-mm-dd