Mini Review

Advantages and advancements of multiple imputation

Abstract

Multiple imputation is still an underused approach for handling missing data despite new advances and its potential in clinical, environmental, and health policy research. This review will discuss several benefits of this technique as well as the approaches that make the technique applicable to different types of data. Providing such examples may show investigators how considering this method may help in their future research.

Volume 2 Issue 3 - 2015

Irene B Helenowski
Department of Preventive Medicine, Northwestern University, USA

Correspondence: Irene B Helenowski, Department of Preventive Medicine, Northwestern University, USA, Tel (312) 503-3597, Fax (312) 908-9588,
Email i-helenowski@northwestern.edu

Received: April 14, 2015 | **Published:** April 16, 2015

Introduction

Multiple imputation expands the possibilities of different analyses involving complex models which would otherwise not converge given unbalanced data caused by missingness. An example of this scenario involved linear mixed effects models with repeated measures (Lindstrom and Bates, 1989; Milliken and Johnson, 1992). In such models, parameter estimates are commonly obtained by the restricted maximum likelihood (REML) algorithm. The computation involved with this algorithm cannot estimate the numerous parameters included in the model such as the within-subject variation, however, where covariates exhibit different patterns and amounts of missingness. This situation can be remedied through imputation where parameter estimates can be obtained from each imputed, balanced data set and averaged for each parameter.[1-3]

Creating new avenues of analyses without collecting further data would be beneficial in terms of cost. Investigators may determine how to pursue their objectives given the significance of associations in their imputed data. Such an approach would aid in studies where each data point could be difficult and expensive to obtain.[4] Analyses on the imputed data may be used to aid in choosing which variables provide the most insight into the questions proposed by the study at hand.

Multiple imputation has also been developed with consideration of the missing data mechanism, a facet potentially ignored with other approaches. This mechanisms including missing-completely-at-random (MCAR), missing-at-random (MAR), and non-ignorable missingness. Under the MCAR mechanism, missingness is independent of the observed and missing data and under the MAR mechanism, missingness is only dependent on observed data.[1,5] Multiple imputation techniques have also been developed for the non-ignorable mechanism where missingness depends also depends on the missing data.[6-8] Demirtas[6] describes one approach of proceeding with imputing non-ignorable missing data using a pattern mixture model and incorporating indicator variables for the dropout groups.

Methods have also been developed for handling missing data which are non-normally distributed. Helenowski & Demirtas,[9] Helenowski et al.,[10] and Helenowski & Demirtas[11] discuss imputing non-normally distributed continuous data, binary data, and mixed non-normally distributed continuous data and binary data, respectively, allowing for the relaxation of assumptions associated with joint modeling. Here joint modeling involves the normal distribution for imputing continuous data, the multinomial data for imputing categorical data, and the general location model for imputing the mixed data with continuous and categorical variables. Helenowski, Demirtas, and McGee[12] extends the concepts of Helenowski et al.,[10] and Helenowski & Demirtas[11] to imputed mixed data with variables not normally distributed and categorical variables having more than two levels. These approaches involve transforming the data into normally distributed values, applying multiple imputation via joint modeling under assumptions of the normal model and back-transform the imputed values onto the scale of the original data.

Given these examples and review of new multiple imputation approaches presented, this article aims to persuade investigators to consider this technique in their work as its benefits could lead to enhancements in their objectives.

Acknowledgement

None.

Conflict of Interest

No conflict of interest

References

1. Schafer JL. *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London. 1997.
2. Schafer JL, Olsen MK. Multiple Imputation for Multivariate Missing–Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*. 1998;33(4):545–571.
3. Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed models with missing values. *Journal of Computational and Graphical Statistics*. 2002;11(2):437–457.
4. Jovanovic BD, Subramanian, Helenowski IB, et al. Clinical Trial Laboratory Data Nested Within Subject: Components of Variance, Sample Size and Cost. Forth coming in Biometrics and Biostatistics International Journal. 2015.
5. Demirtas H. Modelling Incomplete Longitudinal Methods. *Journal of Modern Applied Statistical Methods*. 2004;3(2):305–321.
6. Demirtas H. Multiple Imputation under Bayesianly Smoothed Pattern–mixture Models for Non–ignorable Drop–out. *Stat Med*. 2005;24(15):2345–2363.

7. Siddique J, Belin TR. Using an Approximate Bayesian Bootstrap to Multiply Impute Non ignorable Missing Data. *Comput Stat Data Anal*. 2008;53(2):405–415.

8. Siddique J, Harel O, Crespi CM. Addressing Missing Data Mechanism Uncertainty using Multiple–Model Multiple Imputation: Application to a Longitudinal Clinical Trial. *Ann Appl Stat*. 2012;6(4):1814–1837.

9. Helenowski IB, Demirtas H. A semi–parametric approach for imputing mixed data. *Statistics and Its Interface*. 2013;6(3):399–412.

10. Helenowski IB, Demirtas H, Erdogan BD. On imputing binary data via pair wise associations and corresponding conditional probabilities. *Turkish Clinics Journal of Biostatistics*. 2012;4(1):1–9.

11. Helenowski IB, Demirtas H. Multiple imputation for continuous data via a semi parametric probability integral transformation. J Biopharm Stat. 2014;24(2):359–377.

12. Helenowski IB, Demirtas H, McGee MF. A semi–parametric approach to impute mixed continuous and categorical data. Health Services and Outcomes Research Methodology. 2014;14(4):183–193.

Citation: Helenowski IB. Advantages and advancements of multiple imputation. *Biom Biostat Int J.* 2015;2(3):93–94. DOI: 10.15406/bbij.2015.02.00033