

Multiple Imputation of binary outcomes in small sample size scenarios –A comparative study of Predictive Mean Matching and Bayesian Logistic Regression

Thursday 24 July 2025 11:45 (15 minutes)

Poster

Multiple Imputation of binary outcomes in small sample size scenarios –A comparative study of Predictive Mean Matching and Bayesian Logistic Regression

Author

Ermioni Athanasiadi

Affiliation

University of Siegen

Abstract

Missing binary outcomes in randomised controlled trials commonly pose challenges for accurate statistical inference. Multiple Imputation (MI) has emerged as a widely used approach to address missing values under Missing At Random Mechanisms (MAR) and has been found to be superior to other methods in the context of logistic regression [3]. However, MI methods have their limitations when sample sizes are (very) small [1]. Although MI generally outperforms other missing data handling techniques in small samples [2], little research has investigated these consequences in the binary outcome context, where the required sample size further depends on the imbalance of event rates in the outcome.

This study aims to systematically investigate the performance of two common multiple imputation methods –predictive mean matching (PMM) and bayesian logistic regression imputation (BLR) –in handling missing data for binary outcomes in scenarios characterized by small sample sizes.

Using Monte Carlo simulations ($n = 1,000$) and a 2-level factorial design, a comparative study was conducted. The performance of both imputation methods was evaluated in terms of unbiasedness of estimates, computational efficiency (runtime in R), and feasibility (convergence performance of imputation algorithm). For each method, four parameters were manipulated: amount of missingness (10%, 20%, 40%, 60%), sample size (20, 50, 100, 200, 500), size of regression coefficient (small/medium) and the amount of zeros in outcomes (medium/low). Multiple imputation was applied using the pmm and logreg methods from R package mice [4]. The findings revealed that while both methods produce unbiased estimates for sample sizes greater than 200, significant bias emerges for smaller samples, particularly with PMM yielding larger standard errors. The runtime between both methods is comparable, although PMM shows some advantages. Notably, convergence issues arise more frequently for imputations with PMM and increase with smaller sample size, high missingness rates, and higher imbalance of outcome events. The findings indicate that while both methods perform well under small sample sizes up to 200, important limitations arise for sample sizes of 200 downwards.

The findings are discussed with regard to the applicability of both PMM and BLR to provide guidance for methodological considerations of researchers, when they have to handle missing binary outcomes in small sample scenarios.

[1] Kleinke, Kristian (2018): Multiple Imputation by Predictive Mean Matching When Sample Size Is Small. In: Methodology 14 (1), S. 3–15. DOI: 10.1027/1614-2241/a000141.

[2] McNeish, Daniel (2017): Missing data methods for arbitrary missingness with small samples. In: Journal of Applied Statistics 44 (1), S. 24–39. DOI: 10.1080/02664763.2016.1158246.

[3] Meeyai, S. (2016). Logistic Regression with missing data: a comparisson of handling methods, and effects of percent missing values. Journal of Traffic and Logistics Engineering, 4(2).

[4] Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of statistical software, 45, 1-67.

Keywords

multiple imputation, missing, binary, pmm

Primary author: ATHANASIADI, Ermioni (University of Siegen)

Co-author: Dr KLEINKE, Kristian (University of Siegen)

Session Classification: Poster Session 3

Track Classification: Statistical analyses: Statistical analyses