

MULTIPLE IMPUTATION TECHNIQUE: HANDLING MISSING DATA IN REAL WORLD HEALTH CARE RESEARCH

Rapeepong Suphanchaimat^{1,2}, Supon Limwattananon^{1,3} and Weerasak Putthasri¹

¹International Health Policy Program (IHPP), Ministry of Public Health, Nonthaburi;
²Banphai Hospital, Ban Phai; ³Faculty of Pharmaceutical Sciences, Khon Kaen University,
Khon Kaen, Thailand

Abstract. The problem of missing data is ubiquitous in real world health care research. Its consequences include an introduction of bias and a loss of statistical power. Several methods to account for missing data are proposed. In recent literature, Multiple Imputation (MI) has become more widely used. This study therefore sought to exemplify and discuss the pros and cons of the MI application in relation to some conventional methods, which were Mean Substitution and Regression Imputation. Data from previously published article were used. The variables of interest were serum creatinine and age. Counterfactual dataset with missing data was generated and used to assess to what extent each method estimated the missing values close to the actual data. Under Missing At Random (MAR) assumption, the study presumed that age was related to the missingness of serum creatinine. Mean Substitution tended to produce biased estimate than other methods. Regression Imputation, although producing less biased estimate, did not account for data uncertainty. MI yielded mean and standard deviation estimates closest to the actual data, compared to other two methods. To sum up, MI seems to have advantages over Mean Substitution and Regression Imputation, as it can preserve data variability and also account for uncertainty of missing data. However, it is worth noting that, more importantly than which method is used, the researcher(s) should check thoroughly if the missingness mechanism is not missing not at random (MNAR); and such assessment needs firm background knowledge on the aims/objectives and the methodology of the study.

Keywords: age, introduction of bias, loss of statistical power, Mean Substitution, missingness mechanisms, Multiple Imputation, Regression Imputation, serum creatinine, Tenofovir

Correspondence: Rapeepong Suphanchaimat,
International Health Policy Program (IHPP),
Ministry of Public Health, Nonthaburi 11000,
Thailand.

Tel: +66 (0) 2590 2366; Fax: +66 (0) 2590 2385
E-mail: rapeepong@ihpp.thaigov.net