The Performance of Multiple Imputation in Social Surveys with Missing Data from Planned Missingness and Item Nonresponse

Julian Axenfeld

University of Mannheim

Designs using planned missingness, such as the split questionnaire design, are becoming more and more important in social survey research. To ensure an acceptable questionnaire length, these approaches typically entail large amounts of planned missing data, which can be imputed after data collection. However, social surveys typically also include other types of missingness such as item nonresponse by survey participants, which need to be imputed as well. This entails a complex imputation task with amounts of missing data larger than initially planned and a potentially non-random, heterogeneous mechanism. Since previous research regarding the imputation of planned missingness did not take additional nonresponse into account, it remains to be studied whether accurate multiple-imputation estimates can be obtained in practice with planned missingness and item nonresponse.

To deal with this research gap, in this paper we apply a Monte Carlo simulation study using real social survey data from the German Internet Panel. In this study, we simulate missing data based on item nonresponse with different mechanisms and proportions of item nonresponse as well as different proportions of planned missing data. We find that item nonresponse can jeopardize the quality of estimates after multiple imputation especially when the total amount of missing data from both sources is high or when there is a considerable proportion of item nonresponse that is missing not at random. Therefore, from an imputation perspective, survey designers should incorporate their expectations about item nonresponse on each variable when designing surveys with planned missing data.