Are trajectories of dataset representativeness during survey data collection generalizable? Evidence from the 2011 Census Non-Response Link Study.

Jamie Moore*
Administrative Data Research Centre for England and Department of Social Statistics and Demography, University of Southampton, Southampton, UK. Email: j.c.moore@soton.ac.uk
Gabriele Durrant
Administrative Data Research Centre for England and Department of Social Statistics and Demography, University of Southampton, Southampton, UK. Email: g.durrant@soton.ac.uk
Peter W.F. Smith
Administrative Data Research Centre for England and Department of Social Statistics and Demography, University of Southampton, Southampton, UK. Email: p.w.smith@soton.ac.uk

Abstract

Methodologists have moved away from using survey response rates to indicate risks of non-response bias. Instead, indicators estimating dataset representativeness in terms of variation in sample frame response propensities have been developed. Incorporating statistical models predicting response propensities, and so requiring attribute information on the entire sample frame, these indicators also allow variation to be decomposed to assess the impact of respondent response propensity variation associated with multiple attribute variables. Coupled with call record paradata, they can also be used to monitor data collection over time, and so can inform adaptive strategies in which methods may be modified between fieldwork periods with the aim of maximizing data quality / minimizing costs. Use of these indicators is growing, but gaps in our knowledge remain, especially regarding adaptive strategies. For example, the generalizability of trajectories of dataset representativeness over data collection across different surveys of the same sample frame is unstudied. Hence, it is unknown whether rules for ending data collection with minimal effect on representativeness derived for one survey of a population are applicable to another. We address this question using sample frame census attribute data and call record paradata for two UK social surveys. We use R indicators and Coefficients of Variation of response propensities to describe dataset representativeness trajectories over collection for each survey, and consider whether it is possible to end collection early without affecting data quality. We then compare trajectories in each survey, and evaluate whether rules are generalizable.

Keywords: Non-response bias, representativeness indicators, adaptive data collection, paradata.