

“Driven to the max? Explaining mode-specific survey response propensity using the most ‘ideal’ auxiliary variables we could think of and measure”

Thomas Klausch (Utrecht University)

Barry Schouten (Statistics Netherlands)

Joop Hox (Utrecht University)

Survey modes cause selection processes in household and person surveys to differ. In this paper we seek to understand and model such differences as well as possible. Analyses of selection processes are normally constrained by available information, which often do not exceed national registries. However thinking about what type of ‘ideal’ auxiliary variable one would need for our purpose (and a given survey) one encounters two major questions, (a) ‘What are the ‘ideal’ variables to explain nonresponse in general, for a particular survey, and between modes?’, and (b) ‘If there is an ‘ideal set’, how can we measure it for all population units?’.

With regards to (a) we figure that while approaches from general sociological, economic and psychological theories may help to guide the selection of promising explanatory concepts, there is also considerable reason to believe that a share of response variation is closely associated with the survey’s target variables (or its ‘topic’). Indicators of the latter type may be, for example, just the target variables themselves in the case of an NMAR selection process. The earlier type may be extremely diverse. We hence decided to classify explanatory concepts into three types, registry-type socio-demographics (X), general variables based on relevant behavioural theory (Z), and the target variables or close correlates (Y).

Assuming availability of information on X , we decided to solve problem (b) by a unimode post-hoc survey, which is attached to a regular survey project, in our case the Dutch Security Monitor. Following a first parallel multimode experimental wave (CAPI, CATI, CAWI, PAPI; $n=8800$) where respondents were randomly assigned to one mode only, we re-approached both respondents and nonrespondents in a second wave (unimode CAPI, i.e. the mode with highest response expectation) to gather information on Z and Y . Y were a subset of variables from wave 1 and Z contained measures on topic involvement, general survey attitudes, and attitudes about the survey sponsor. We will use X , Z , and Y to model mode-specific response propensities in nested models and seek to maximize fit. We hope to find cross-validated subsets of (X , Z , Y) that fully explain between-mode selection processes.