Undercoverage and Bias in Name-based Samples of Foreigners

Rainer Schnell and Tobias Gramlich (University of Duisburg-Essen, Duisburg, Germany),
Mark Trappmann (Institute for Employment Research (IAB), Nuremberg, Germany)

In many cases, there are no sampling frames for rare or special populations, like foreigners or migrants. Therefore, often sampling frames for a more general population are (automatically) screened for members of the target population. Whereas this is an efficient way of sampling rather rare and specific populations, knowledge of potential limitations and threats of this widely used approach is very limited.

(Automatically) screening and classifying members of a population according to some criteria may produce false positive matches (e.g. natives wrongly classified as foreigners) as well as false negatives (e.g. foreigners wrongly classified as domestic). Whereas false positives only increase screening or survey costs, false negatives potentially introduce bias if they are systematically different in variables relevant to the topic of the survey.

We use a large scale German panel survey ('PASS') on labour market and social security to investigate coverage problems and potential bias when using a Bayesian-based classification of names to screen for foreigners in a general population sampling frame, and to show the trade-off between efficiency and potential bias due to false negatives. We will also show bias in variables relevant to migration or labour market research, introduced by false negative classifications.