# Using auxiliary information in data collection and in estimation

**Kaur Lumiste**

*University of Tartu, Estonia, kaur.lumiste@ut.ee*

**Keywords**: non-response, responsive design, balanced response

High nonresponse is forcing national statistics offices around the world to find new ways of designing and controlling the data collection in their surveys. Responsive design is a newly emerged view focusing on the possibilities to reduce the effects of nonresponse by monitoring the data collection process. The main survey goodness measure thus far – nonresponse rate – is by itself an insufficient guide, since ordinarily collected sets of respondents tend to be biased towards certain society's groups like elderly or people living in rural areas. So more informative measures of the progress of the data collection have recently been proposed like balance indicators ([1]) and R-indicators ([2]) (where R stands for representativeness). In current work we use the first approach and aspire to representativeness through balance of the response set with respect to a given set of auxiliary variables - means of auxiliary variables have to be approximately the same in the sample and the response set.

The same auxiliary variables can also be used at the estimation stage to improve our estimates, but assume that we have access to more auxiliary variables at the estimation stage than we did in the data collection stage. Is the effect of additional explanation power affected by balancing? Finding an answer to this question brings another one - should we emphasise on acquiring more auxiliary variables for the estimation stage or should we focus more on balancing the response? Which would have a larger effect on the bias and/or accuracy of the final estimates? Applying both strategies can be very costly, so budgets can be optimised by opting for the better strategy.

In this paper we look for evidence to support one of the following strategies:

- put effort into collecting more auxiliary information and focus on post-weighting correction;
- put effort into monitoring response to get a representative set of respondents.

Interactions of response propensities, post-weighting weights, auxiliary variables and estimated variable are studied in the context of these strategies.

## References

[1] Särndal, C.-E. (2011). The 2010 Morris Hansen lecture dealing with survey nonresponse in data collection, in estimation. *Journal of Official Statistics* **27**, 1–21.

[2] Schouten, B., Cobben, F., Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology* **35**, 101–113.