Representativity Indicators for Survey Quality

Representativity Indicators for Survey Quality (RISQ) – publishable summary

Barry Schouten
Statistics Netherlands

10th of September 2010

The RISQ Project is financed by the 7th Framework Programme (FP7) of the European Union.
Cooperation Programma, Socio-economic Sciences and the Humanities, Provision for Underlying Statistics

Project goals and background

Project RISQ (Representativity Indicators for Survey Quality) started in 2008 with three goals in mind: to develop and test quality indicators that 1) enable evaluation and comparison of the representativity of response to surveys and registers, 2) can be used as tools in monitoring response during survey data collection, and 3) can be used as objective functions in improving representativity of response through tailored survey designs.
At the start of the project basic ideas had been developed of what quality indicators may look like and how they may be estimated. However, the indicators were incomplete, and above all needed international comparison and empirical validation. The project structure and format were directed at extension and elaboration of existing theory and at international validation of indicators using datasets from five countries.

Quality of response is traditionally measured by the response rate; the percentage of sampled units that responds to the survey. Although, response rates do give valuable insight into the quality of a survey, it does not present a complete picture. Examples of surveys with high response rates but low quality exist in the literature. Representativity of response is a quality that complements response rates, but is rarely made rigourous in statistical literature. The need for definitions of representative response and indicators that measure a lack of representative response were the main motivation for the project.
In order to be really informative and useful, indicators need to have a number of pre-specified properties. If the goal is to compare and evaluate response, then indicators must be easy to interpret, easy to compute and easy to translate to practical settings. If the goal is to monitor survey data collection, then indicators must be relevant and must allow for a top-down assessment of the impact of various population characteristics and survey design features. If the goal is to re-design survey data collection in order to improve representativity, then indicators must be effective.

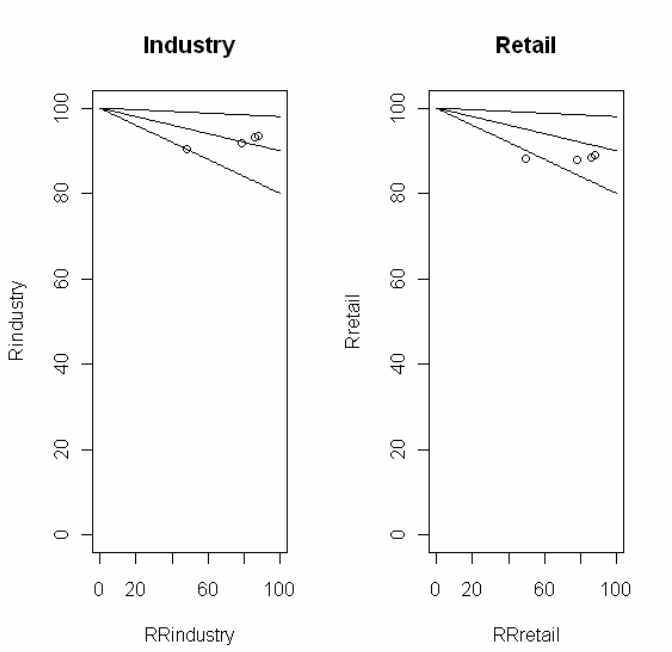
With these goals and these requisites in mind, the project was divided in a theoretical part and an empirical part. The empirical part consisted of two designed data collection experiments.

Project achievements

As it's major achievement RISQ has developed Representativity indicators; simply termed as R-indicators and partial R-indicators. The indicators are documented, extensively tested and implemented as open source modules for R and SAS.
The analyses, tests and experiments conducted within the project show that the indicators are valuable tools and fit for purpose in assessing the quality of survey response and in monitoring and improving the design of survey data collection. Survey designs tailored with the help of the indicators improved the representativity of response significantly. At workshops and conferences the project has advocated the use of the indicators as standards in quality reports. Besides in evaluating and designing survey data collection, the indicators also proved useful in monitoring the completion of registry data.

The R-indicator is proposed as the quality indicator for evaluating and comparing the response to different surveys and to single surveys in time. It's values range between 0 and 1, where 1 corresponds to a perfect response and 0 to the least ideal setting where the impact of nonresponse is maximal. The indicator is based on the variation in response propensities. As such it has a clear interpretation and can be linked to the maximal impact of response on survey statistics. The computation of the indicator is straightforward. However, since adjustments are needed when survey sample sizes are small, it is recommended to use the indicators implemented in R and SAS.

The R-indicator is also the starting point for monitoring and re-designing data collection. Partial R-indicators break down the R-indicator into components related to specific background characteristics like age and business type. The indicators can be computed at the variable level, thus showing the impact of a specific characteristic, say again age or business type, on a lack of representative response. They can also be computed at the category level, for instance persons of 25 years and younger or retail businesses, thereby presenting the impact of specific subpopulations. This property allows for a top-down approach in monitoring response. When R-indicators are low, then partial indicators can be used to zoom in on the impact of specific characteristics and subpopulations.



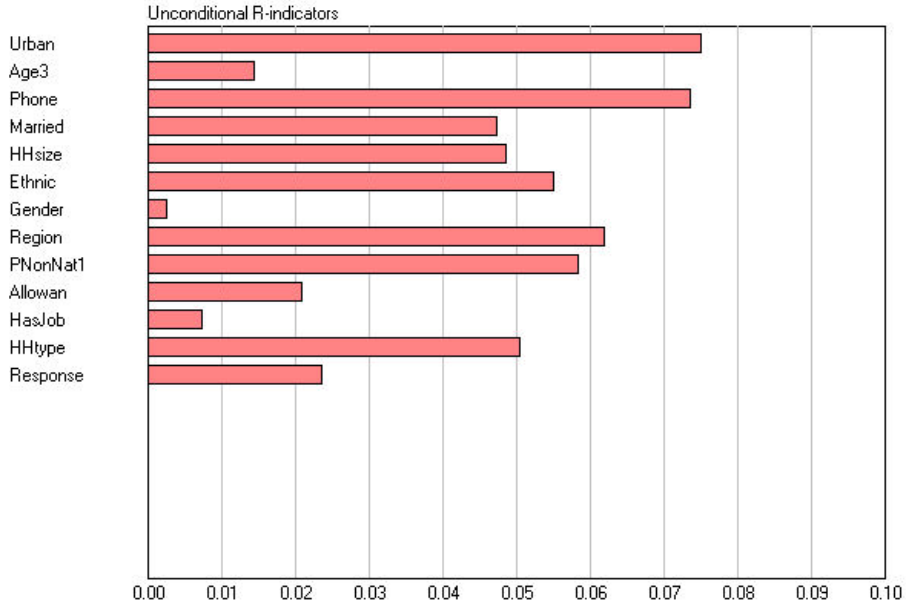
In the recent survey methodology literature survey designs that target subpopulations and tailor survey design features to these groups have become a new paradigm. Such designs are usually termed responsive or adaptive survey designs. In these designs the R-indicators and partial R-indicators can be used as quality objective functions in trade offs with costs. For this purpose they are already tested at various survey institutes. The project tested two such designs and concluded that the indicators are indeed promising tools. In the coming years the indicators will form the input to subsequent research and empirical analysis in this area.

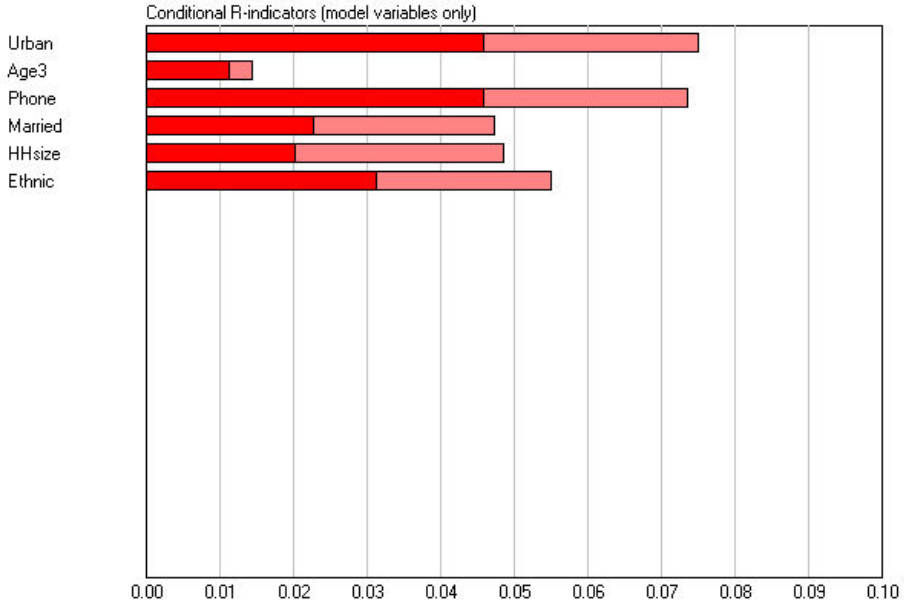
All project papers and products are available at the website www.risq-project.eu. As a guide to the project deliverables:

- Comparisons of response quality over countries or over time – deliverables 2.1, 2.2, 3 and 4
- Monitoring survey data collection – deliverables 5 and 6
- Re-designing data collection – deliverables 7 and 8
- Technical reviews of RISQ results – deliverables 15, 16 and 17
- Software manuals – deliverable 12.1 and 12.2

Two overview papers summarise and discuss the various project results. Deliverable 9, Representativeness indicators for measuring and enhancing the composition of survey response, gives an overview of the theory and concepts behind representative response and the various indicators. Deliverable 10, Monitoring and changing data collection through R-indicators and partial R-indicators, provides an overview of the application of the indicators as well as guidelines for their use in practical settings.

The indicators are implemented in SAS and R open source modules. They can be downloaded from the website. A manual and test data set are available. Next to these modules a graphical tool has been developed to quickly provide insight into the representativity of response. This tool is also available at the website.





Project partners and the future

RISQ is a collaboration of the National Statistical Institutes of Norway (Statistisk Sentralbyrå), Slovenia (Statistični Urad Republike Slovenije) and The Netherlands (Centraal Bureau voor de Statistiek), and the universities of Leuven (Belgium) and Southampton (UK). The project is coordinated by Centraal Bureau voor de Statistiek.

Formally RISQ ended July 2010 with the release of software in SAS and R and the publication of theoretical and practical overview papers. Informal the project will be continued. In November 2010 release 2 of the software is planned that will add standard errors to partial R-indicators. In the spring of 2011 release 3 will include population-based R-indicators. The RISQ website, www.risq-project.eu, will be updated on a regular basis and will contain the future releases. RISQ related questions and a request to receive future announcements can be send to risq@cbs.nl .

Contact persons for RISQ are:

- Centraal Bureau voor de Statistiek: Barry Schouten (overall coordinator, bstn@cbs.nl) and Jelke Bethlehem (jbtm@cbs.nl)
- University of Southampton: Natalie Shlomo (N.Shlomo@soton.ac.uk) and Chris Skinner (C.J.Skinner@soton.ac.uk)
- Statistisk Sentralbyrå: Øyvin Kleven (Oyvin.Kleven@ssb.no)
- Katholieke Universiteit Leuven: Geert Loosveldt (Geert.Loosveldt@soc.kuleuven.be)
- Statistični Urad Republike Slovenije: Katja Rutar (Katja.Rutar@gov.si)