Representativity Indicators for Survey Quality

Q2010 – R-indicators and Fieldwork Monitoring (draft)

Koen Beullens & Geert Loosveldt
Katholieke Universiteit Leuven

May 4th, 2010

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

Optimizing survey efforts and R-indicator through adaptive sample selection

In this paper, an attempt is made to inform the effect of additional fieldwork efforts on the quality of the obtained sample. Response rates will always increase because of renewed contact attempts, but as not always all sample cases receive additional visits, the composition of the obtained sample can be affected by the selection strategy for renewed attempts. We will first deal with a fictitious situation, followed by a handling of a real survey.

In both applications we will assume the 'continuum of resistance' model (Lin & Schaeffer, 1995). This model considers response rates as a function of fieldwork efforts. The more efforts that are invested in the fieldwork, the higher the response rate will be. In a more sophisticated approach, one can distinguish between contact rates and cooperation as two underlying and substantive dimensions in the response process. Augmented fieldwork efforts, however, do not necessarily lead to better sample quality in terms of the difference between respondents and nonrespondents (e.g. Groves, 2006). In this respect, noncontact or refusal conversion may only result in 'more of the same kind of respondents'. This paper seeks to sort out the consequences of three selection strategies, based on initially estimated response propensities: the high propensity selection strategy, the random selection strategy and the low propensity selection strategy. In the two applications, selection strategies are assumed to decide upon during the fieldwork, implying a dynamic or adaptive fieldwork strategy (Groves & Heeringa, 2005).

A simple simulation study

Suppose a survey has to be carried out. The gross sample contains 10.000 elements and interviewers are told only to attempt each case once. During the visit (independent of its outcome) the interviewers observe some of the characteristics of the houses (multi-unit or single unit, overall quality score of the dwelling, observations about the area, etc.). Also the postal codes of all sample cases are known so that auxiliary information can be enriched with municipality-level information (population density, income per capita, crime rates, etc.). We also assume that the sample register contains data about the target persons (e.g. age and gender).

For simplicity, all these auxiliary variables are rolled up into one auxiliary vector X with mean zero and a standard deviation of one. After the first attempt, that requires an effort of 10.000 personal visits, about 50% of the cases have become respondents. The individual response propensities can be expressed as:

$$ln\left(\frac{response = 1}{response = 0}\right) = 0 + 0.5x \, .$$

Now, an additional budget permits to revisit 2.500 cases once in order to improve the quality of the obtained sample. The fieldwork management decides to select those cases independently of the auxiliary information (= random selection among initial nonrespondents). Conversion success is believed to be expressed as follows:

$$ln\left(\frac{conversion = 1}{conversion = 0}\right) = -1 + 0.5x \, .$$

We believe the intercept in the conversion model is somewhat lower as the status of an initial nonrespondent is predictive for the response propensity at a renewed attempt. We do not have any expectation with respect to a possible shift of the effect of X on the conversion success. This situation has been simulated and the sample quality indicators are given in the next table before and after the additional efforts.

Table 1: Sample quality indicators before and after revisits (simulation - 50 replication)

	Before	After
Response rate	0.5012	0.5646
R-indicator	0.7674	0.7540
Maximal absolute contrast	0.4652	0.5003
Maximal absolute bias	0.2321	0.2178
Mean resp. prop among nonrespondents	0.4744	
Mean resp. prop among 2500 selected nonrespondents	0.4741	

Given the specifications about the resampling among the nonrespondents and the logistic regression model assumptions, the investment of an additional 25% contact efforts realizes an increase of the response rate of 6.46 percent points. This improvement however, is discouraged by the mild relapse of both the representativity indicator and the maximal absolute contrast. Therefore, there is only a modest improvement of the maximal absolute bias. The last two rows of the table indicate that the average response propensities among the initial nonrespondents is slightly lower then the overall sample propensity average. The mean propensity of the selected nonrespondents is very close the average of all nonrespondents, indicating that the selection of 2.500 units from the nonrespondents is representative of all initial nonrespondents.

We believe that the small maximal absolute bias improvement can not justify the additional 25% contact efforts. Particularly, the blind re-selection of cases generates only 'more of the same'. High propensity cases that were initially not responding will more easily be converted than their low propensity equivalents. This only reinforces the already existing gap between respondents and nonrespondents (unless the parameter for X would have been smaller in the conversion phase that in the initial phase; however, it is hard to find a reason why such a decrease would take place).

Provided that the (1) selection of the 2.500 cases is the strategic asset upon which the management can autonomously decide and that (2) ultimately the maximal absolute bias is the quality indicator that is to be minimized, we will simulate two supplementary selection scenarios. The first scenario only selects the 2.500 highest propensity cases to be reissued. Here, it is expected that the response rate will strongly increase, although the diversity of response propensities will strongly grow, resulting in an even worse maximal absolute bias. Only the 2.500 lowest propensity cases will be selected in the second scenario. Under these conditions, the response rate is expected to grow slower that in the first scenario, whereas the final propensities after reissuing will converge. This will force the maximal absolute bias to improve considerably.

Table 2: Sample quality indicators before and after revisits, 3 scenarios (simulation - 50 replication)

	Initial phase	High prop. selection	Random selection	Low prop. selection
Response rate	0.5012	0.5841	0.5646	0.5424
R-indicator	0.7674	0.6575	0.7540	0.8466
Maximal absolute contrast	0.4652	0.7051	0.5003	0.3091
Maximal absolute bias	0.2321	0.2932	0.2178	0.1414
Mean resp. prop among nonrespondents	0.4744			
Mean resp. prop among 2500 selected nonrespondents		0.5658	0.4741	0.3747

The columns 'Initial phase' and 'Random selection' have been copied from the first table. The highest and lowest propensity selection scenarios have been added. After 50 simulations of the two scenarios, our expectations seem to be supported. High propensity selection results in a stronger increase of the response rate, as compared to the random selection and the low propensity selection. The variability of response propensities is also stronger in the high propensity selection, indicated by the R-indicator. The contrast between respondents and nonrespondents is much smaller after low propensity selection.

The best return-on-investment seems to result from the selection of the lowest propensity cases. Under high propensity selection, it seems the sample quality (maximal absolute bias) will strongly deteriorate, despite the increase of the response rate.

This simulation illustrates that the same amount of efforts lead to different quality arrangements, depending on the selection strategy. If one seeks to maximize the response rate, high propensity selection should obviously be considered, whereas the selection of low propensity cases makes the propensities converge. This latter strategy should however be considered with care. It can be imagined that because of an over-selection of low propensity cases, an inverse situation emerges: high propensity profiles become underrepresented and vice versa. A second concern addresses, as always, to the quality of the auxiliary information. In the example we assumed that the available auxiliary information is relevant to inform about the damage to the obtained sample caused by nonresponse. Only as long as the available set of auxiliary variables is a fair approximation of $\aleph$, the inferences drawn from the quality indicator after the different selection strategies are valid. This latter assumption is of course hard to verify.

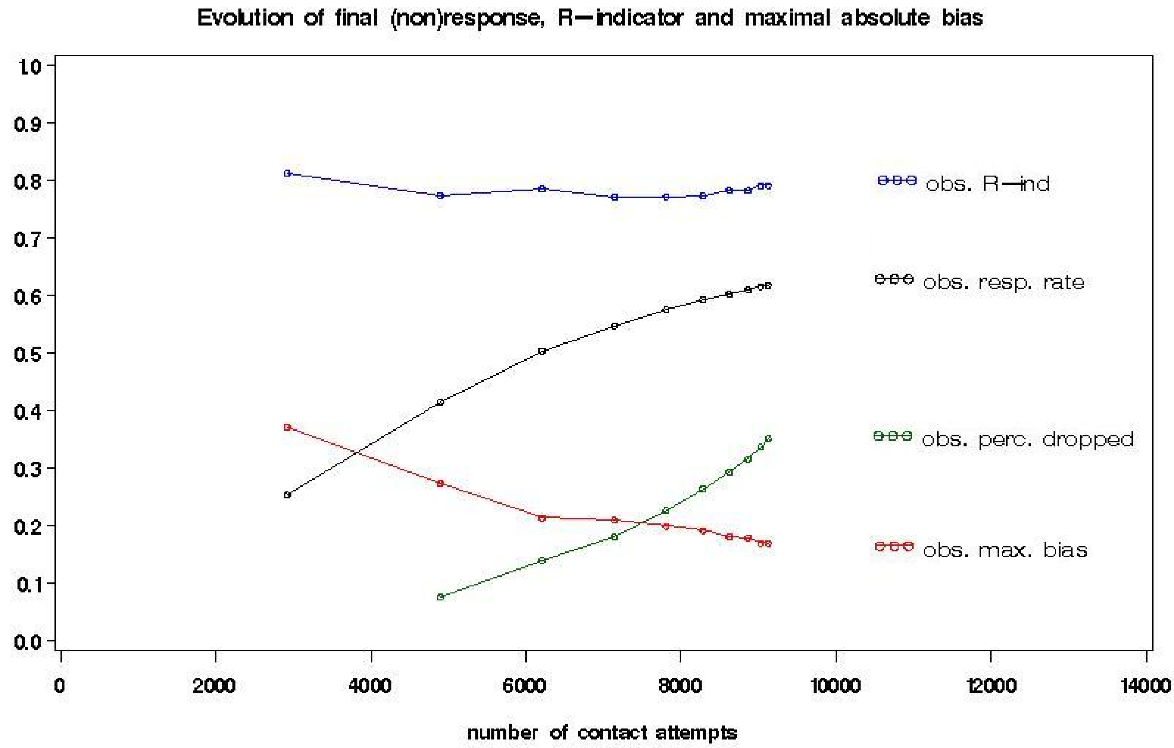
Empirical illustration: ESS3 – BE

The gross sample of the Belgian part of the ESS counts 2927 cases. 1829 of them have finally been converted into respondents. For some units, only one attempt was enough, others needed more fieldwork efforts. This means that for 1089 cases no additional contact efforts have been done, although they did not comply with the (several) survey request. Some of the nonrespondents were not reissued even after the first unsuccessful attempt, other were re-approached several times. The next table provides the information about how many cases were converted into respondents or were finally considered as nonrespondents per contact attempt. After the first attempt, 740 cases had been interviewed, 219 were never contacted again (despite the unsuccessful attempt) and 1968 were still pending. Those 1968 units were reissued, of which 476 were successful, 189 were never tried again, etc.

Table 3: Evolution of respondents and nonrespondents per contact attempt. ESS3 - BE

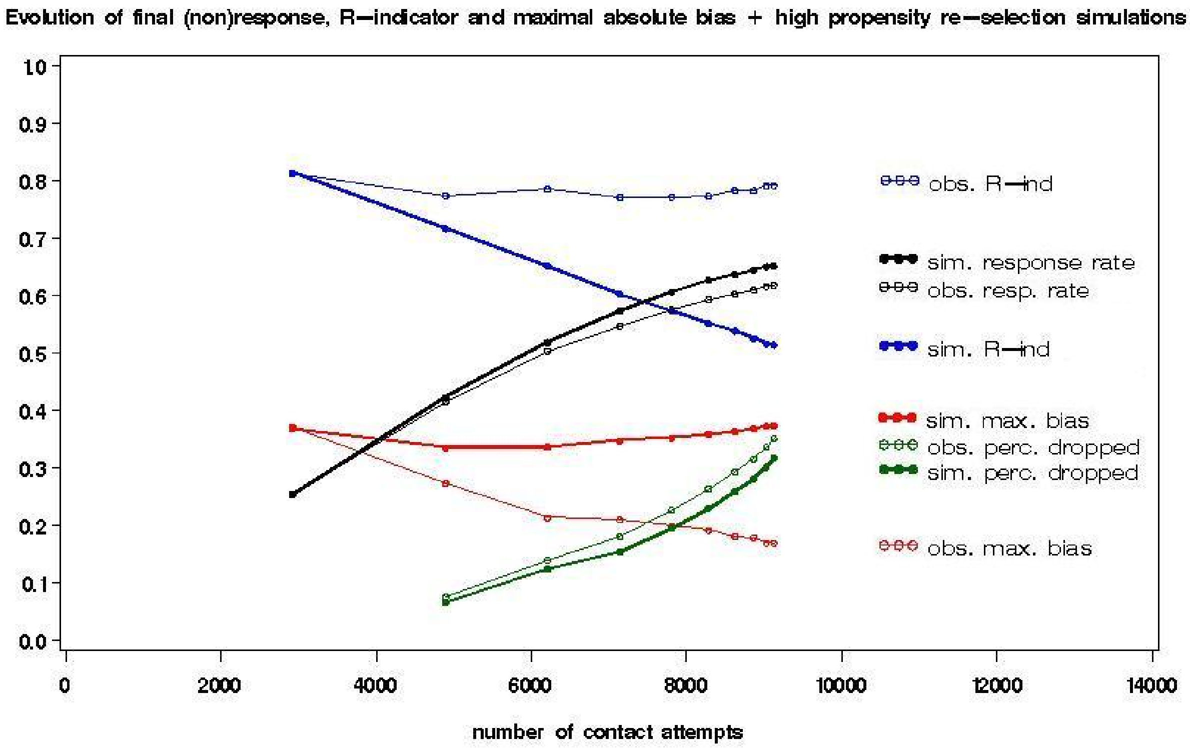
Attempt	Respondents	Nonrespondents	Pending	Respondents (cumulative)	Nonrespondents (cumulative)
0	0	0	2927	0	0
1	740	219	1968	740	219
2	476	189	1303	1216	408
3	263	119	921	1479	527
4	131	136	654	1610	663
5	84	109	461	1694	772
6	51	84	326	1745	856
7	28	68	230	1773	924
8	22	59	149	1795	983
9	19	44	86	1814	1027
10	5	39	42	1819	1066
11	3	14	25	1822	1080
12	5	9	11	1827	1089
13	2	9	0	1829	1098

This table shows clearly that at any time, sample cases are abandoned, while others are re-selected for additional fieldwork efforts. We will investigate what the proceeds are of those additional fieldwork efforts. Therefore, sample quality indicators will be determined after each contact attempts. Thereafter, the two selection scenarios will be simulated. In these simulations, all fieldwork conditions will be held constant, such as the number of contact attempts, the number of selected / abandoned cases at each contact attempt, the success probability conditional on the auxiliary information per contact attempt. Only the selection probability will be controlled. After each contact attempt based on the propensities at the first contact attempt, the highest / lowest propensity cases will be selected. As auxiliary information we have the following variables at our disposal: age, gender, type of dwelling (single unit versus multi-unit), quality of the dwelling as observed by the visiting interviewer and area information such as the population density, average income of the municipality, the percentage of foreigners living in the municipality and the region in which Belgium the sample unit resides (Flanders, Wallonia or Brussels). The first graph shows the actual situation and can be used as a reference.



The black line indicates the evolution of the response rate as a function of the number of attempts. After the first contact attempt that counts for 2927 efforts, about 25% if the sample has been converted into respondents. After the second contact attempt 4895 (= 2927 + 1968) efforts have been done resulting in a response rate of 42%. Finally after 13 contact attempts, counting for about 9.500 contact efforts, the response rate is 62%. The green line shows how the percentage of the units finally considered as nonrespondents grows. The blue line shows what happens with the variability of the response propensities during the fieldwork process, as expressed by the R-indicator. It seems that the diversity of propensities hardly changes, if not deteriorates. Only at the end of the fieldwork process, a small recovery emerges. Finally, the maximal absolute bias steadily improves (red line), probably and exclusively due to the increase of the response rate.

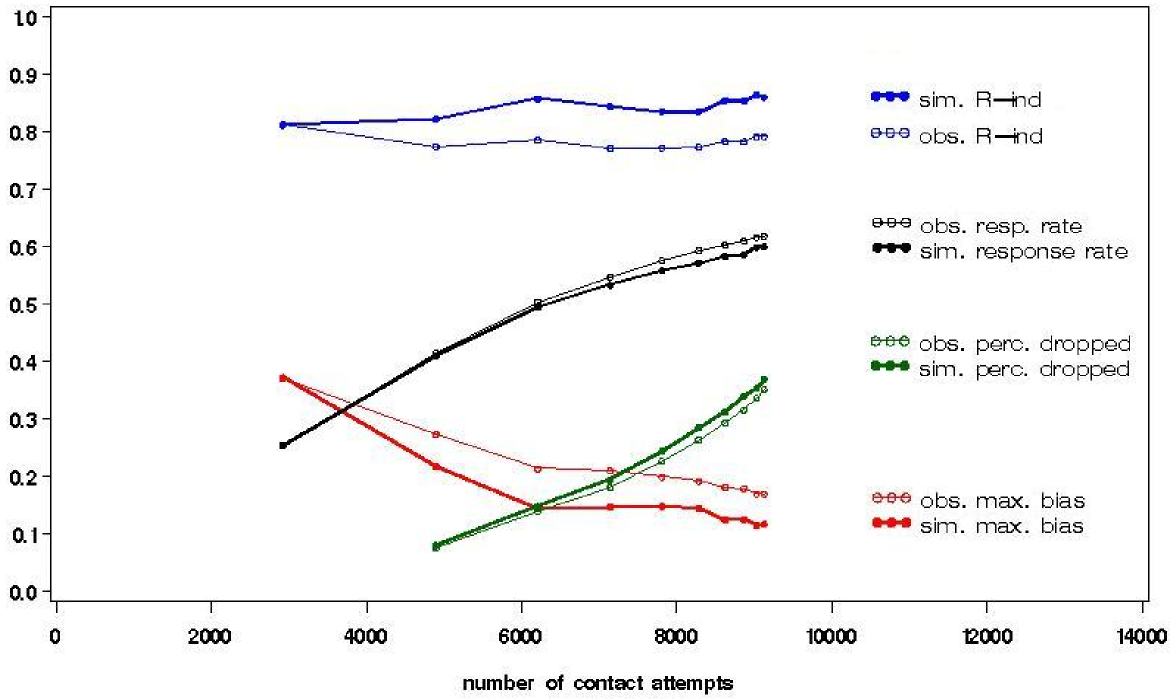
The next graph shows, in addition to the actual situation, the simulated sample quality indicators provided that consistently the high propensity cases have been selected.



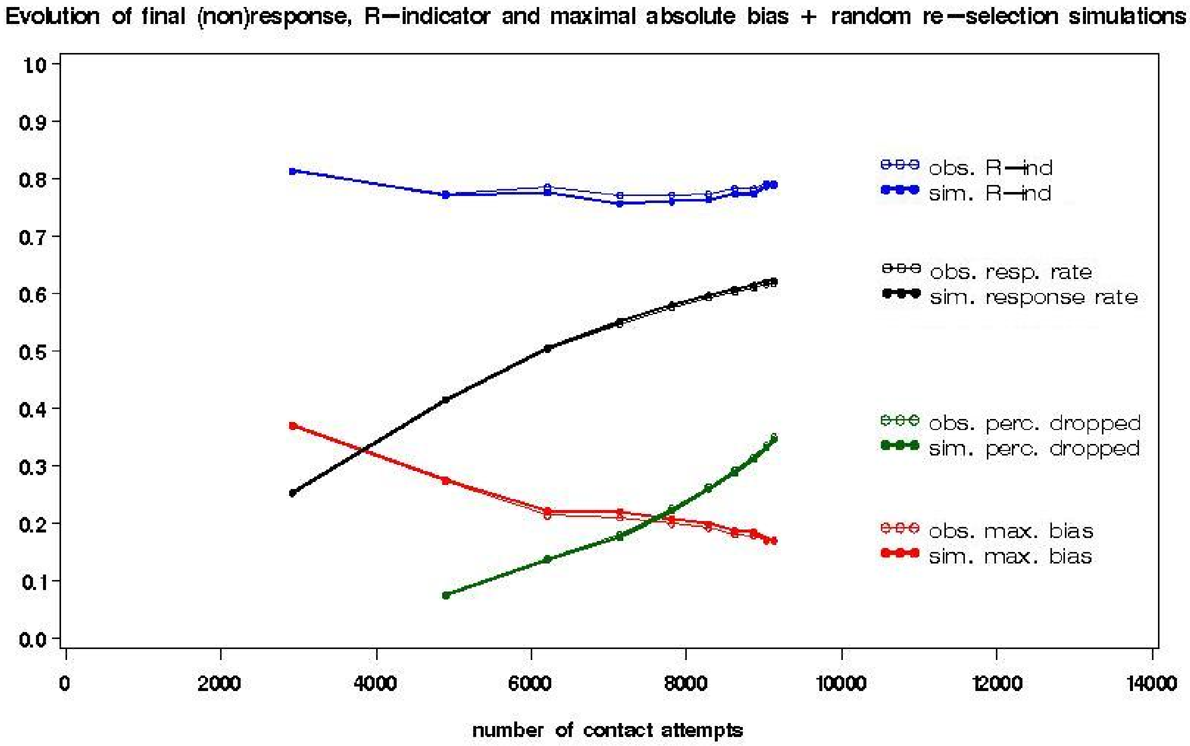
Provided that consistently high propensity cases are selected, the diversity of propensities steadily grows, as indicated by the R-indicator (thick blue line). Compared to the thin blue line (observed R-indicator) the representativity decreases substantially. The response rate, as expected, grows faster compared to the reference evolution. The maximal absolute bias estimate seems to slightly worsen as the number of efforts increase. On the condition that the bias estimate is the most important quality indicator, all additional contact efforts prove to be a waste of energy, under this selection strategy.

The next scenario shows completely different results. Here, the lowest propensity cases have consistently been selected for renewed attempts. The response rate increases not as fast as compared to the real fieldwork, and grows much slower than the high propensity selection regime. On the other hand, the R-indicator turns out to evolve much more prosperously, as well as compared to the high propensity scenario as to the observed situation. As a result, the tentative estimation of the absolute maximal bias is much more advantageous than in the previous two situations. Given that the actual fieldwork needed about 9.500 efforts to achieve a maximal absolute bias of 0.17, the same sample quality can be obtained after only less then 6.000 (linear interpolation) efforts when consistently low propensity cases have been prioritized for re-selection.

Evolution of final (non)response, R–indicator and maximal absolute bias + low propensity re–selection simulations



Finally we will also consider the situation in which random selection would have taken place. It is expected that the actual fieldwork has proceeded this way. The reference lines (indicating the actual situation) can hardly be seen in the graph above. They coincide very strongly with the simulated lines, where random selection was instructed. This makes the assumption of blind re-selection most likely.



Discussion

This attempt to inform the possible consequences of purposive selection of sample cases for reissuing suggests that more efficiency can be attained from strategic fieldwork decisions. Giving priority to low propensity cases on the one hand results in better sample quality indicators (except the response rate), or on the other hand, the same selection strategy generates similar results as blind (random) selection but needs less fieldwork efforts. As far as fieldwork efforts indicate the cost of a survey, purposive selection can make surveys cheaper. This purposive selection however presupposes a good set of auxiliary information for the estimation of the propensities. However, as every sample case usually needs be visited once (even without making contact), interviewers' eyes can be used to collect auxiliary information such as litter and rubbish lying around in the neighbourhood, the condition of the dwelling, etc. Furthermore, area information can be linked to the postal codes. Register variables such as age and gender may also be useful auxiliary information.

A second concern points to the attention for response rate in contemporary survey research. Response rate have in recent years been criticized for being poor indicators of sample quality. Particularly the high propensity selection simulation supports this concern and urges to focus more on alternative quality indicators such as the R-indicator or the estimation of maximal absolute bias.

This exercise in selection strategies is of course only explorative. More research efforts on the limitations or pitfalls of selection strategies are most welcome.

References

Lin, F., & N. Schaeffer (1995) Using survey participants to estimate the impact of nonparticipation. Public Opinion Quarterly, 59(2), pp. 236-258

Groves, R., & S., Heeringa (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society: Series A (Statistics in Society), 169(3), pp. 439-457.

Groves, R. (2006) Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70(5), pp. 646-675.