RISQ

Representativity Indicators for Survey Quality

Indicators and data collection control
Report from the pilots in Statistics Norway and Statistics Netherlands

Work package 7, Deliverable 7

Johan Fosen, *Øyvin Kleven* *and* *Bengt Oscar Lagerstrøm* *Statistics Norway*	*Annemieke Luiten,* *and* *Willem Wetzels* *Statistics Netherlands*

September 2010

SEVENTH FRAMEWORK PROGRAMME

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

Acknowledgements
Thanks to Einar Bjørshol and Li-Chun Zhang for their valuable contribution.

Contents

1. Introduction

The impact of nonresponse on survey quality is typically measured by the response rate. The response rate alone, however, is not sufficient as a quality indicator to capture the potential impact of nonresponse. The bias of estimates resulting from nonresponse also depends on the contrast between respondents and nonrespondents with respect to a target variable. The more they differ, the larger the bias will be. Since the target variable is only known for the respondents, we need indicators that measure the degree to which the group of respondents resembles the complete sample with respect to variables available in the whole sample. Such indicators are currently lacking. The RISQ project (Representativeness Indicators for Survey Quality, www.risq-project.eu) was set up in order to find such indicators that measure the representativeness of the response. We call these indicators *Representativity indicators* or *R-indicators*.

The indicators can be used in different settings, e.g.,

- To compare the response to different surveys that share the same target population, e.g. households or businesses
- To compare the response of consecutive waves of a repeated survey .
- To monitor the response to a survey during data collection, e.g. after various days, weeks or months of fieldwork
- To control the response to a survey by means of adaptive survey designs or responsive survey designs (e.g. Groves and Heeringa 2006, Mohl and Laflamme 2007, Wagner 2008)

R-indicators can serve as quality indicators in two stages of the production of statistics. Firstly they can be used as *Product quality,* as in the first two examples above. On the other hand, the two last examples show us the r-indicators used for *Process quality*, which will be focus in this paper. When process quality is the aim, R-indicators will be used as a monitoring device to aid the surveillance of surveys during fieldwork. R-indicators can give information on which survey to focus on and partial R-indicators can give information on to which sub sample we should allocate more and less resources. Both nationally and internationally, we have seen a growing focus on how to improve response rates, minimize response bias and monitor fieldwork costs during the fieldwork period. Former strategies have been to make specifications of and to standardize all aspects of design, then to implement those specifications and finally to analyse conditional on these design protocols. However, the probability of obtaining an interview with a sample member cannot be fully predicted, and this type of uncertainty unfortunately implies lack of control over both the cost, timeliness, and ultimately the error structure of our survey data.

A solution to this uncertainty is to have a flexible survey design that is adapted to the survey process. The fixed design features that formerly have been specified prior to the initiation of data collection, include sample design, sample size, length of data collection period, number of interviewer hours, travel costs, and, sometimes, number of calls to cases prior to the first contact, number of contacts, type and level of calling following a sample person expressing some reluctance to participate. In the presence of the uncertainties, some of these may be candidates for real time data collection alteration. Combining easily retrieved administrative data and process data make continuous monitoring of survey variables of interest possible. The task is to maximize the result, given certain constraints as time or costs. Computer assisted interviewing with a fully integrated computer exchanging system provides rich data on the interviewing process and gives the opportunity to analyse paradata and the data from the survey itself during the fieldwork period (immediately after the survey has started) (Thomsen et al. 2006). Real-time access to information about the survey process enables the survey organisation to analyse the quality of the fieldwork process and to make mid-course decisions and design alterations during the fieldwork period (Hapuarachchi, March and Wronski 1997; Couper

1998; Scheuren 2001). This is often known as "responsive survey design" (Heeringa and Groves 2005), and is the topic of the present paper.

Currently, there are no explicit quality indicators for monitoring and changing data collection. Subgroup response rates come closest to such indicators but are defined only at the category level and do not reflect the relative impact of those subgroups on the overall response rate or response representativeness. Partial R-indicators fill this gap. In this paper we assess their use in monitoring and changing data collection.

Schouten et al., (2010a) describe requirements to indicators used for monitoring data collection: Indicators should be relevant in the sense of pointing to groups that should be targeted, and they should allow for zooming in and out between an overall view and detailed views. Furthermore, the indicators need to be effective: Interventions based on indicators should lead to a more representative response and to less nonresponse error. The current paper describes these features of R-indicators. In two pilots, the relevance and effectiveness of R-indicators were studied. R-indicators are compared to analysis of subgroup response rates to show how each of these measures is able to reflect the overall response quality.

One pilot was held by Statistics Norway, and one by Statistics Netherlands. In both, a differential fieldwork design was used, in which different strategies were assigned to different groups in the sample. The decision which groups to target was aided by partial R-indicators. In the pilot of Statistics Norway this is done during fieldwork, in a truly responsive design. In the pilot of Statistics Netherlands, the differentiation occurred before the commencement of fieldwork, based on prior knowledge of comparable sample units' behaviour in similar surveys.

This paper starts with a short overview of R-indicators and partial R-indicators. This introduction is followed by the setup, results and discussion of the Norwegian pilot. The third part describes the Dutch pilot. This part is followed by an overall discussion.

2. R-indicators and Partial R-indicators.

In this Section, we will briefly describe and define representativeness, before we present the R-indicators. Most of this section is an adaption of Schouten & Shlomo (2010b). See Schouten et al. (2009), Shlomo et al. (2009a), and Shlomo et al. (2009b) for details.

2.1 The concept of Representativeness

As in Schouten et al. (2009), we define a response set to be representative if the individual response probabilities are equal for all units in the population. Because individual response probabilities are impossible to estimate based on a single response for each sample unit, we restrict ourselves to response propensities. We will let X denote auxiliary variables in the survey, e.g. register variables or data covering the whole sample from other sources. We let ρ_X denote the response propensity function for variable X, say the vector age and gender. This means that $\rho_X(x)$ is the probability of response for a population unit with $X = x$, say young females.. We suppose that X is a subset of a vector of auxiliary variables that explains response behaviour and for which the response propensities can be viewed as individual response probabilities.

We use the two definitions of Schouten et al. (2009) for representative response and conditional representative response respectively.

Definition: A response to a survey is representative with respect to X when response propensities are constant for X, i.e. when $\rho_X(x)$ is a constant function.

Definition: A response to a survey is conditional representative with respect to X given Z when conditional response propensities given Z are constant for X, i.e. when $\rho_{X,Z}(x,z) = \rho_Z(z)$ for all x.

As described in the introduction, since the target variable or vector of target variables Y is unknown for the nonrespondents, we confine ourselves to representativity with respect to X. Preferably X should be highly correlated with Y, but also X should be chosen in order to allow comparisons in the different settings described in the introduction. We define our main target as answering the question whether data collection succeeded in obtaining a balanced response for a set of pre-selected variables X that is available before or during data collection.

2.2 Measuring deviations from representative response

For measures the distance between two vectors of response propensities ρ_1 and ρ_2, we use the Euclidean distance

$$d(\rho_1, \rho_2) = \sqrt{\frac{1}{N} \sum\nolimits_U (\rho_{1,i} - \rho_{2,i})^2}\,, \tag{2.1}$$

where N is the population size, U the population and i the unit.

Using (2.1), the R-indicator is defined as a transformed distance between all the N individual response propensities ρ_X in the population, and the individual responses under representative response. The latter is the N-dimensional vector $\rho_0 = (\rho, \rho, \ldots, \rho)^T$, where ρ is the survey response. The R-indicator is

$$R(X) = 1 - 2d(\rho_X, \rho_0) = 1 - 2S(\rho_X)\,. \tag{2.2}$$

Here $S(\rho_X)$ is the standard deviation of the individual response propensities, which we by (2.1) recognise as the distance $d(\rho_X, \rho_0)$. The transformation ensures that R is bounded by one and zero, where the former means a representative response, and the latter indicates the largest possible deviation from representative response.

Typically, X will be a vector of auxiliary variables like age, gender or urbanization for household surveys and business type and size for business surveys. If measuring representativeness is restricted to one auxiliary variable, say Z, then we have a partial representativeness indicator or partial R-indicator. At the variable level the partial R-indicator is

$$P_u(Z) = d(\rho_Z, \rho) = S(\rho_Z)\,, \tag{2.3}$$

the standard deviation of the response propensity function $\rho_Z(z)$ in the population. The subscript u in (2.3) indicates that this is an partial indicator for unconditional representative response, as opposed to partial R-indicators for conditional representative response that we will define in section 2.3.

For any Z it holds that $P_u(Z) \in [0,1]$. Furthermore, $P_u(Z) \in [0, (1 - R(X))/2]$ when Z is an element of X.

For categorical variables we also define partial R-indicators for each category in addition to the variable-level indicator. Assume that Z have $k = 1,2,\ldots,K$ categories and let ρ_{Z_k} response propensity function for the group where $Z = k$, for example that age is less that 35 years old. The partial R-indicator for a category k is defined as

$$P_u(Z,k) = \sqrt{\frac{N_k}{N}}(\rho_{Z_k} - \rho), \tag{2.4}$$

where $N_k = \sum_U Z_k$ is the number of population units in category k . Negative values indicate under-representation of category k while positive values indicate over-representation, and $P_u(Z,k) \in [-1,1]$. We see that $P_u(Z,k)$ is a function of the response function only for the individuals where $Z = k$, and we notice that the variable level indicator is related to the category level indicator by

$$P_u(Z) = \sqrt{\sum_{k=1}^{K} P_u^2(Z,k)} ,$$

2.3 Measuring deviations from conditional representative response

For measuring representativeness of one variable while controlling for the impact of other variables, we have a conditional partial R-indicator,

$$P_c(Z \mid X) = d(\rho_{X,Z}, \rho_X) = \sqrt{\frac{1}{N-1}\sum_U (\rho_{X,Z}(x_i, z_i) - \rho_X(x_i))^2} , \tag{2.5}$$

the distance between the individual propensities based on both X and Z , and those based only on X . If for example X is household composition, household income and province of residence while Z equals the age of the main person in the household, $\rho_{X,Z}(x_i, z_i)$ are the propensities based on all these variables, and $\rho_X(x_i)$ only based on the age.

By restricting the sum in (2.5) to the part taken over class $Z = k$, we have the conditional partial R-indicator on category level

$$P_c(Z,k \mid X) = \sqrt{\frac{1}{N-1}\sum_U Z_k (\rho_{X,Z}(x_i, z_i) - \rho_X(x_i))^2} . \tag{2.6}$$

We cannot assign a positive or negative sign to the category level conditional partial indicators, indicating over- or under-representativeness. The reason is that the sign may be different for each subclass of X . For some combinations of household composition, income and province, we may experience that age category k has a positive effect on response while in others it has a negative effect.

It turns out that (2.5) is the square root of the “within-X" variance of the $\rho_{X,Z}$ propensities, i.e. the variation of the $\rho_{X,Z}$ propensities within the categories of X.. . In our example, it represents the variation in response behaviour due to the age of the main household person given both its household composition, income and the province of the household.

3. Introduction to the Pilot of Statistics Norway

The purpose of the pilot in Norway was to gain experience with using the R-indicators and to evaluate the usability of R-indicators in survey fieldwork. For the test group a responsive design was used, where the design alterations during fieldwork were based on monitoring using R-indicators. For the control group, the standard procedure of Statistics Norway was used. This includes monitoring of response propensities ρ_X where X is univariate or bivariate, and reassignments of units to interviewers based on the monitoring results. In Section 3.1 we will look more closely into how monitoring has been used in Statistics Norway, before we return to the pilot in Section 3.2.

3.1 Monitoring during fieldwork at Statistics Norway

Monitoring and responsive design is not new in Statistics Norway and below we will say something about the current status in Statistics Norway. In the handbook Coping with decreasing response rates in Statistics Norway Recommended practice for reducing the effect of nonresponse (Thomsen et al 2006) it was proposed to monitor the development in response rate and response bias during the fieldwork. This was proposed to be visualised by tools that are easy to make and interpret. The rationale behind the tool was that the variables used as explanatory variables in the presentation of the estimates, as well as the variables used in post stratification and weighting adjustments, should guide the decision on whether to stop or continue the fieldwork. This means that some variables must be considered more relevant in terms of representativity than others. In attempts to reduce nonresponse bias during fieldwork, tables produced by simple cross tabulations of some important background variables were used. There is a long tradition to monitor auxiliary variables (e.g. variables from registers where we know the value for all the elements in the sampling frame) separately during the fieldwork process. Figure 1 gives an example from The Election Survey 2005, and here Gender is displayed as the percentage of over or under representation of Males. If the distribution of males and females are the same in the gross sample as in the net sample, the value will be 0. We see that males are overrepresented by ca 3 percentage points in the beginning of the fieldwork, after 10 weeks the bias is slightly decreasing and drops to about 2 percentage point. For almost all the variables the bias decreases as we go along in the fieldwork.

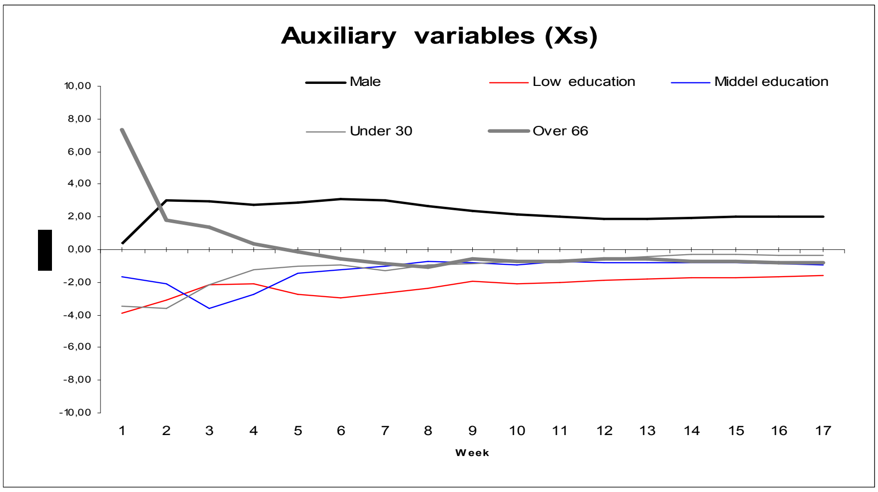

Figure 1. Development of some auxiliary variables during fieldwork.

Simple tools like the one shown in figure 1 are of course useful but we want a tool that can combine all the variables in figure 1. Otherwise we are in danger of introducing more bias in one variable if we decide to increase our effort to recruit more women into the net sample. Table 1 shows an example where we have combined all the variables in Figure 1. The gross sample is stratified by the combinations of three register variables from the sampling frame, Gender (male/female), Age (young/middle/old) and Education (low/middle/high). This gives a matrix containing 18 cells, for which the difference between response proportion in the gross sample and the net sample is calculated every week. If the difference exceeds 0.5 percent the cell turns blue if the difference is negative and red if it is positive.

Table 1. Difference in percentages between gross sample and net sample in the Election Survey 2005 by week in fieldwork. Gender, Age group and Education.

		Male									Female								
Age		Young	M	Old	Young	M	Old	Young	M	Old	Young	M	Old	Young	M	Old	Young	M	Old
Education		Low			Middle			High			Low			Middle			High		
Week	**1**	-0,3	-1,6	1,0	0,2	-3,2	2,0	-1,7	1,4	1,9	-0,2	-1,8	-1,0	-2,3	0,7	0,8	1,0	0,6	2,4
Week	**2**	-0,5	-0,6	0,6	-0,4	-0,9	1,1	-0,5	2,8	1,0	-0,7	-0,8	-1,2	-2,0	0,6	-0,5	0,6	0,7	0,6
Week	**3**	-0,1	-0,3	0,3	-0,3	-1,0	1,1	0,5	2,0	0,7	-0,7	-0,3	-1,1	-1,7	-1,4	-0,3	0,3	1,8	0,5
Week	**4**	0,0	-0,2	0,0	-0,1	-0,7	0,8	0,3	1,9	0,7	-0,5	-0,2	-1,2	-1,2	-1,3	-0,2	0,4	1,3	0,3
Week	**5**	-0,1	-0,5	-0,2	0,2	0,1	0,6	0,3	1,9	0,5	-0,5	-0,5	-1,1	-1,1	-1,1	-0,2	0,3	1,2	0,2
Week	**6**	-0,2	-0,2	-0,2	0,3	0,6	0,3	0,2	1,9	0,4	-0,5	-0,8	-1,1	-0,9	-1,2	-0,3	0,4	1,1	0,2
Week	**7**	-0,2	0,0	-0,3	0,0	0,8	0,3	0,2	1,7	0,3	-0,5	-0,6	-1,1	-0,9	-1,1	-0,2	0,2	1,2	0,1
Week	**8**	-0,2	0,2	-0,2	0,2	0,6	0,2	0,2	1,4	0,1	-0,5	-0,6	-1,2	-0,7	-0,8	-0,3	0,1	1,2	0,1
Week	**9**	-0,2	0,2	-0,1	0,3	0,2	0,3	0,1	1,3	0,1	-0,4	-0,5	-0,9	-0,6	-0,9	-0,2	0,0	1,0	0,1
Week	**10**	-0,2	0,2	-0,1	0,3	0,3	0,3	0,1	1,1	0,1	-0,5	-0,6	-0,9	-0,4	-1,1	-0,3	0,1	1,5	0,0
Week	**11**	-0,2	0,1	0,0	0,3	0,4	0,2	0,1	1,0	0,1	-0,4	-0,6	-0,9	-0,3	-0,9	-0,3	0,0	1,4	0,1
Week	**12**	-0,1	0,0	0,1	0,1	0,4	0,2	0,2	1,0	0,1	-0,3	-0,6	-0,9	-0,3	-0,9	-0,2	0,0	1,3	0,1
Week	**13**	-0,1	0,0	0,1	0,1	0,5	0,1	0,1	1,0	0,1	-0,3	-0,6	-0,8	-0,3	-1,0	-0,2	0,1	1,2	0,1
Week	**14**	-0,1	0,1	0,0	0,2	0,4	0,1	0,2	1,0	0,0	-0,2	-0,7	-0,8	-0,3	-1,0	-0,2	0,1	1,1	0,1
Week	**15**	-0,1	0,0	0,0	0,1	0,4	0,1	0,2	1,1	0,0	-0,2	-0,7	-0,8	-0,3	-1,0	-0,2	0,1	1,1	0,1
Week	**16**	-0,1	0,0	0,0	0,1	0,5	0,1	0,2	1,1	0,0	-0,2	-0,6	-0,8	-0,3	-1,0	-0,2	0,1	1,1	0,1
Week	**17**	-0,1	0,0	0,0	0,1	0,5	0,1	0,2	1,1	0,0	-0,2	-0,6	-0,8	-0,3	-1,0	-0,2	0,1	1,1	0,1

One obvious advantage of using table 1 is that it is pretty simple to make and interpret for fieldwork managers. On the other hand, even with only three variables it becomes quite difficult to follow. If we include more variables it will be over-complex, and also the number of persons in some cells will often be very small when many variables are combined. The latter can be solved by looking at coefficients from a logistic regression model (without too many interaction terms). However, in most survey organisations, employees working as managers and supervisors of the interviewer process do not have sufficient background in statistics to handle such models.

In our early approach we suggested to look at the Chi square of the divergence between the gross sample and the net sample for every week, if the Chi square of the distribution would be lower, then the 'representativity' could be said to have improved (Thomsen et al. 2006:26). It is evident that this approach needed to be further developed and the overall R-indicator is a better and even easier model to interpret from the point of view of a statistician. And it was also evident for us in the survey organisation that we needed a tool, to monitor the response portion in each sub stratum of interest, that was more sophisticated and where we could include more variables than the example displayed in Table 1. On the other hand it was also evident that this tool had to be practical to use in a natural survey setting. But what about the practician and survey manager who actually are going to use this in their practical work, will it be any use for them?

3.2 Method

The aim of the pilot at Statistics Norway was to test how R-Indicators and Partial R-indicators would work as tools to facilitate differentiated fieldwork strategies during data collection. We investigate how to differentiate strategies and how the indicators can play a role in the corresponding choices. We will also discuss the benefits of R-indicators in light of other tools for enhancing the Composition of Survey Response used by Statistics Norway

As the pilot for implementing and testing R-indicators, an ongoing survey was selected: the Level of Living Survey 2009 (LLS 2009). An overview of the survey is presented in Table 2. The main purposes of the LLS

2009 are to produce statistics on the work environment in different occupational groups and industries and to analyse this with respect to background characteristics like gender, age, education etc. In order to monitor the development of work environment over time, the LLS 2009 is a panel survey which is repeated every third year, the first round being conducted in 2006. Hence, most of the 20 500 respondents, aged 16 to 69 years, were also contacted three years ago, while a small fraction are new respondents. The survey is primarily conducted by telephone but face-to-face interviewing is permitted. Average interview time was 24 minutes in 2006. The fieldwork period for the survey was from June until November 2009.

Table 2. An overview of the sample for the pilot.

Survey	Survey of living and work environment 2009. A panel survey on work environment subjects which is repeated every third year, the first round being conducted in 2006.
Population	All persons living in Norway between 16-69 years old.
Sample size for the main survey	20 500
Sample design for the main survey	Self-weighting, 2 stage design. Most of the sample of 20 500 where also contacted three years ago, in order to also use the sample as a cross sectional survey a small fraction of young persons and immigrants are included in the sample in each wave.
Data collection mode for the main survey	Mainly telephone but face to face is allowed if a telephone number is not available (ca 10 % of the sampled elements) or if the respondent asks about it.
Interview length	24 minutes in average.
Field work period for the main survey	June – December 2009
Sample design for the pilot	3000 respondents selected randomly from the sample of 20 500. Then the sample of 3000 was randomly split into two equally sized samples of 1500.
Data collection mode for the pilot	Only telephone by interviewers working in the CATI facility in Oslo
Field work period for the pilot	October – November 2009

The fieldwork period for the pilot was set up to take place in October and November 2009. Before the main survey started, we had to make sure that we had a random sample for the pilot. We drew a random sample of 3 000 respondents from the main sample of 20 500, and then randomly split the sample of 3000 into two equally sized groups: a test group for doing interventions based on r-indicators, and a control group where the standard procedure was to be used. The sample of 3000 was not sent into fieldwork before the pilot started in October. For the pilot we only used telephone as interview mode, and we decided to use only interviewers at the CATI facility in Oslo in order to have control over the experiment.

A lot of information is available about characteristics of the sample units from different administrative registers at Statistics Norway. From registers we have access to general demographic information such as gender, age, education, and various geographical variables for all sampled elements. The sample of LLS 2009 was linked to a dataset with some of the variables from Statistics Norway's Population Database. The database is a copy of the Central Population Register[1], which is maintained by the National Tax Administration[2]. The register contains comprehensive information on people living in Norway, such as names, addresses, family size, citizenship, identification numbers, position of employment and marital status. The data is gathered for tax, electoral and population purposes by local tax offices. The Population Database furthermore refines some variables in order to provide more background variables, such as immigration groups, and how central to a city centre a persons dwelling is located. Hence, the Population Database yields three standard demographic background variables to be used in our monitoring, namely age, sex and centrality. The LLS 2009

[1] In Norwegian: "Folkeregisteret".
[2] In Norwegian: "Skattedirektoratet"

sample was also linked to Statistics Norway's Education Register, which contains information on various aspects of the Norwegian population's education. Data on the respondents' level of completed education was imported to monitor the data collection. Survey research shows a strong relationship between level of education and response propensity in household surveys; respondents with a low level of education are less inclined to participate in surveys. Table 3 below shows the univariate distribution in the test sample and the control sample of the background characteristics age (grouped), gender, centrality and level of education.

Table 3. Distribution of background variables in test sample and control sample. Absolute numbers.

	All	Test sample	Control sample
All	3000	1500	1500
Status 2006			
Interview	1 854	934	920
Refusal	402	198	204
Temporary prevented because of illness, work/school or language problems	92	48	44
Non-contact	356	169	187
New respondents - not part of the 2006 survey	296	151	145
Age group			
18-34	944	491	453
35-59	1 553	765	788
60-69	503	244	259
Sex			
Male	1 570	782	788
Female	1430	718	712
Centrality			
Most peripheral	275	135	140
Less peripheral	192	93	99
Less central	503	254	249
Most central	2 030	1 018	1 012
Education			
Primary school and lower secondary school	948	461	487
Upper secondary school	1 207	614	593
Higher education	845	425	420

Setting up a test group and a control group

In figure 2 below we show the standard procedure of the interview process at Statistics Norway. The first two steps, the planning phase (step 1) and the sampling phase (step 2), are common for all surveys at Statistics Norway,. In these two steps, the questions, administrations, modes, data collection period, sample frame, and sampling procedure etc. are decided. Then sample units are assigned directly to interviewers or to the CATI call management system (step 3). In step 4, the interview phase starts: contact and cooperation. In step 5, all sample units with status non-contact or non-cooperative are evaluated, and re-assigned if the sample unit is regarded as possible for re-contact. This procedure is almost completely subjective and not containing considerations regarding representativeness. In our pilot, step 1, 2, 3, and first part of step 4 will be equal for the two groups.. The difference between the standard procedure and the experiment procedure is step 4b, where the experiment group was subjected to close supervision using the R-indicator and partial R-indicators, while the standard procedure was used in the control group. The evaluation lead to different decisions in the re-assignment phase 5.

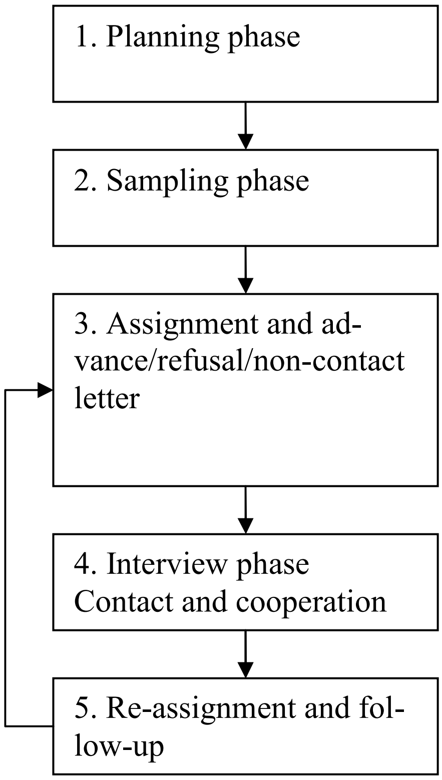

Figure 2. Interview process of the control group

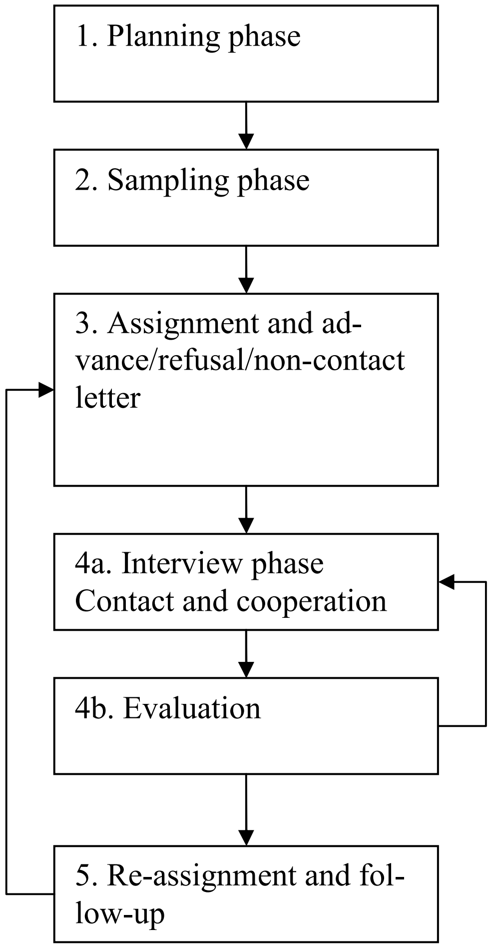

Figure 3. Interview process of the pilot group

Responsive design

At the start, the R-Indicators and partial R-Indicators were registered on a daily basis in order to observe how the indicators fluctuate over time. However, we soon learned that it was more rational to evaluate progress once a week and then decide upon follow-up action. First we had only three variables Age group, Gender, and Centrality in the model, but after some time we expanded it to also include level of Education and Status from the 2006 survey. Table 4 gives an overview of the pilot.

Table 4. An overview of the pilot in Statistics Norway.

	Test group	Control group
Monitoring and supervision	Monitoring by the research group. Cases are re-assigned on the basis of achieving better representativeness. R-indicators calculated throughout the fieldwork, meetings in the project group twice a week.	Regular monitoring by the usual supervisors, and cases is re-assigned without any consideration and priorities regarding representativeness
Interventions during fieldwork	1. Based on R-indicators (3 variables). Prioritise young adults in the CATI call schedule. Use mobile phone numbers instead of land-line 2. Based on 5 variables. Prioritise former non-respondents in the call schedule. Exceeded briefing of selected interviewers on persuasion strategies.	No interventions

Table 5 shows the (unconditional) partial R-indicators by day of the fieldwork, and figure 4 shows the conditional one. The italic values in table 5 represent the variable level unconditional R-indicators, while the other values represent the category level values. If we look at the unconditional partial R-indicator after 18 days for the test group, we see that among the three upper variables that were used at day 18, the deviations from representativity were largest in the age groups, and since the youngest group had the negative value of largest magnitude, this group was most underrepresented.

Table 5. Unconditional Partial R-Indicators for test group and control group. Five explanatory variables [3]

	Test group Days of fieldwork													Control group Days of fieldwork												
	14	16	18	20	22	23	24	25	26	27	28	29	32	14	16	18	20	22	23	24	25	26	27	28	29	32
Age group	*4,4*	*4,5*	*4,2*	*4,4*	*4,0*	*4,2*	*4,3*	*4,6*	*4,6*	*4,5*	*4,2*	*4,4*	*4,1*	*3,1*	*3,2*	*3,8*	*4,0*	*4,0*	*4,3*	*4,1*	*3,8*	*4,0*	*4,1*	*4,2*	*4,3*	*4,2*
18-34	-26	-24	-23	-25	-30	-36	-38	-41	-42	-42	-41	-42	-41	-32	-32	-35	-36	-37	-37	-36	-35	-36	-36	-36	-41	-40
35-59	-13	-15	-14	-13	-7	-2	1	3	4	4	4	5	5	-2	-2	-2	-2	-1	-3	-3	-3	-2	-3	-4	2	1
60-69	59	60	58	60	55	54	53	53	53	52	50	51	49	46	47	50	52	51	54	52	51	51	52	54	51	51
Gender	*0,0*	*0,0*	*0,1*	*0,1*	*0,2*	*0,1*	*0,1*	*0,0*	*0,0*	*0,0*	*0,0*	*0,1*	*0,0*	*0,2*	*0,1*	*0,2*	*0,3*	*0,3*	*0,2*	*0,3*	*0,3*	*0,3*	*0,3*	*0,0*	*0,0*	*0,1*
Male	2	4	8	7	9	7	5	5	5	4	4	5	5	-10	-8	-9	-11	-11	-11	-12	-12	-11	-11	-4	-3	-5
Female	-2	-4	-8	-7	-10	-8	-6	-5	-5	-4	-4	-6	-5	10	8	10	12	12	11	13	12	12	12	4	3	6
Centrality	*0,1*	*0,2*	*0,4*	*0,4*	*0,4*	*0,4*	*0,6*	*0,6*	*0,6*	*0,6*	*0,6*	*0,6*	*0,6*	*0,6*	*0,6*	*0,4*	*0,3*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,1*	*0,2*
Most peripheral	-4	-7	-15	-16	-15	-17	-19	-21	-21	-22	-21	-21	-23	2	1	0	-1	-4	-3	-3	-5	-5	-6	0	-3	-7
Less peripheral	8	12	14	12	10	10	14	12	12	12	10	10	8	-16	-16	-15	-14	-6	-5	-6	-9	-7	-8	-12	-11	-12
Less central	4	3	2	5	2	0	3	3	4	3	3	4	1	18	19	11	9	5	5	4	6	6	6	3	0	3
Most central	-3	-2	0	0	2	3	1	2	2	3	3	3	5	-5	-5	-1	0	1	0	1	2	1	2	2	4	5
Education	*1,8*	*2,2*	*2,5*	*2,5*	*2,9*	*3,6*	*3,6*	*4,1*	*4,2*	*4,1*	*4,1*	*4,2*	*4,5*	*2,7*	*2,9*	*3,8*	*3,9*	*4,0*	*4,1*	*3,9*	*4,7*	*4,8*	*4,6*	*5,0*	*5,8*	*6,0*
Primary school and lower secondary school	-28	-29	-31	-32	-36	-41	-40	-42	-42	-42	-40	-41	-42	-37	-38	-44	-44	-44	-43	-42	-45	-47	-45	-46	-49	-51
Upper secondary school	-2	-5	-5	-4	-2	-1	-3	-4	-4	-4	-7	-6	-7	3	3	4	4	2	0	0	-2	-1	-1	-4	-4	-3
Higher education	32	36	39	38	40	44	45	49	49	49	50	50	52	37	38	43	43	46	47	46	51	51	50	54	58	58
Status in 2006	*14,9*	*14,6*	*16,4*	*17,0*	*18,4*	*19,5*	*18,7*	*20,2*	*20,5*	*20,1*	*19,4*	*19,5*	*19,5*	*17,6*	*17,1*	*19,2*	*19,4*	*20,4*	*21,0*	*20,7*	*22,8*	*22,9*	*23,2*	*25,0*	*27,0*	*27,9*
Interview	68	68	73	74	77	80	79	83	83	82	80	80	81	78	77	82	83	84	86	85	89	90	91	94	98	99
Refusal	-80	-77	-84	-86	-89	-94	-90	-93	-94	-94	-94	-95	-95	-76	-76	-78	-77	-79	-81	-83	-86	-87	-87	-93	-94	-97
Unable	-37	-40	-41	-38	-38	-32	-30	-31	-31	-32	-32	-33	-34	-15	-13	-16	-17	-18	-19	-20	-22	-23	-23	-25	-25	-23
No contact	-50	-49	-47	-49	-53	-55	-55	-58	-58	-55	-52	-51	-49	-71	-69	-76	-77	-81	-80	-77	-81	-80	-81	-80	-87	-88
New respondents – not part of the 2006 survey	-4	-7	-11	-13	-13	-16	-18	-21	-21	-21	-19	-19	-21	-18	-19	-21	-20	-17	-18	-16	-19	-20	-21	-21	-22	-22

*Partial R-indicators * 1000*

The conditional partial R-indicator on variable level at the upper left panel of figure 4 confirms that the under-representation of age is not only a marginal tendency but also holds when we are conditioning on gender

[3] The model with three variables (Age, Gender and Centrality) renders the same unconditional R-indicators as the current table for these variables.

and centrality. Then, as far as we can tell from our three auxiliary variables, age is the reason for under-representation.

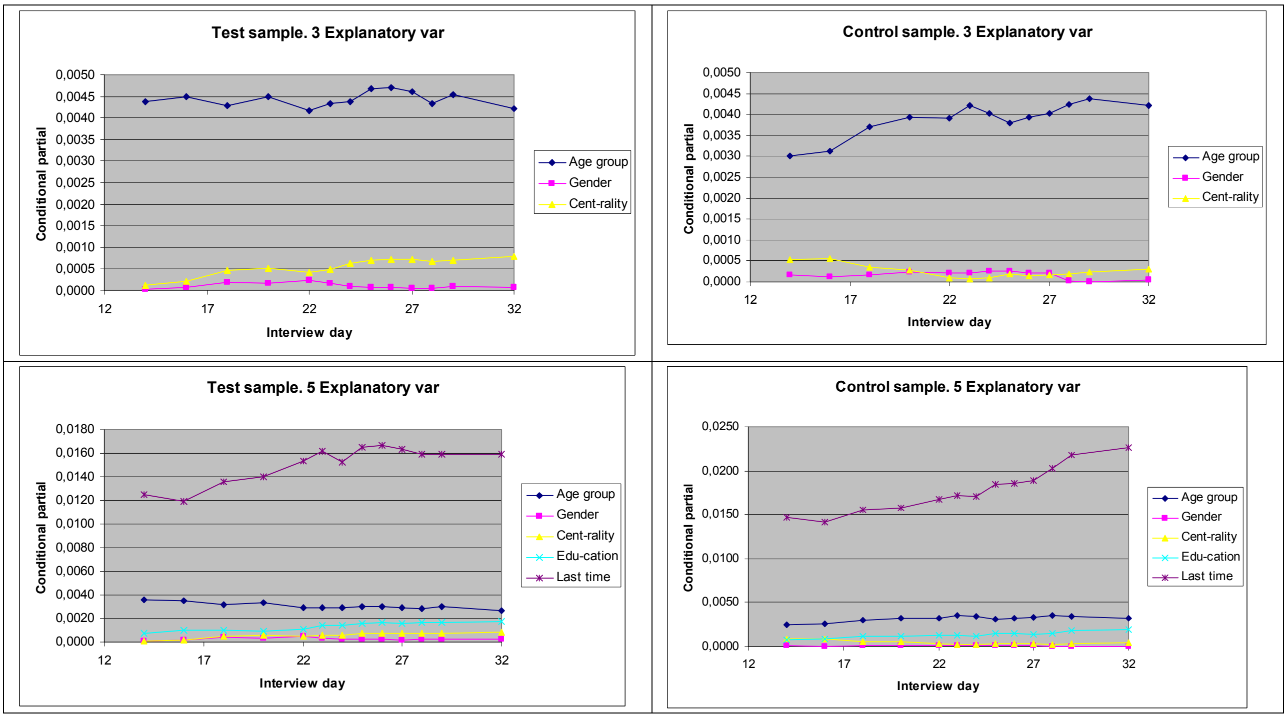

Figure 4. The conditional partial R-indicators on variable level by interview day. Three explanatory variables for test group (upper left panel) and for control group (upper right panel); Five explanatory variables for test group (lower left panel) and for control group (lower right panel) .

Based on the unconditional partial indicator[4], our first intervention was done on day 18 to stimulate the chance of obtaining information from young adults (under 35 years). Our remedy was to prioritise young adults in the CATI call schedule. In addition, our interviewers were told to only use available mobile phone numbers based on a hypothesis that young people are easier to contact and more willing to respond to the survey request when contacted by mobile phone.

The second and last intervention was conducted on day 24, after we had also included status in 2006 and education into our model. In table 5 we see that previous refusers are the group with lowest participation in the survey. Based on results from the partial R-indicators we changed prioritising in the call schedule from young adults to former nonrespondents and the interviewers were briefed on persuasion strategies. Figure 4 confirms that status in 2006 is the most important variable.

3.3. Results - Did we improve representativeness?

In figures 5 and 6, the overall R-indicator for both the test group and the control group is plotted, in order to see the development trough the fieldwork period. The confidence interval for the test sample is also included. We consider figure 5 to be very interesting, since it is close to a chart we will use in the survey organisation of Statistics Norway. Figure 5 is based on a model with age group, gender, centrality, and education as explanatory variables. In figure 6, also status in 2006 is included.

[4] The conditional partial R-indicator on variable level was not available in the output of the SAS-program when the pilot study went on.

In figure 5, the test and control groups are fairly equal until day 27. The small differences before day 27 can not be attributed to any result of the first intervention since the differences between the curves between the day 18 (the day of the first intervention) and day 27 are of approximately the same magnitude as before day 18. After day 27, i.e. three days after the second intervention, there is a somewhat larger difference which gives us some faith that our efforts had a positive effect on the representativity and that it is possible to use the R-indicator system for making quick decisions during fieldwork. Figure 6 gives clear indications that the effort of the second intervention (prioritise former non-respondents in the calling schedule) paid off.

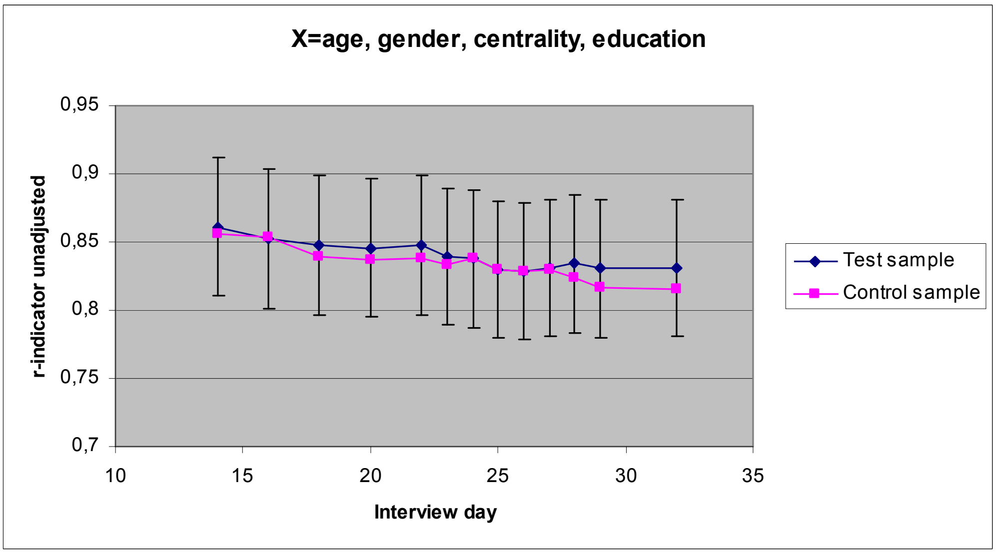

Figure 5 R-indicators for Test sample and Control sample using the four explanatory variables age, gender, centrality and education. Confidence interval for the test sample indicator.

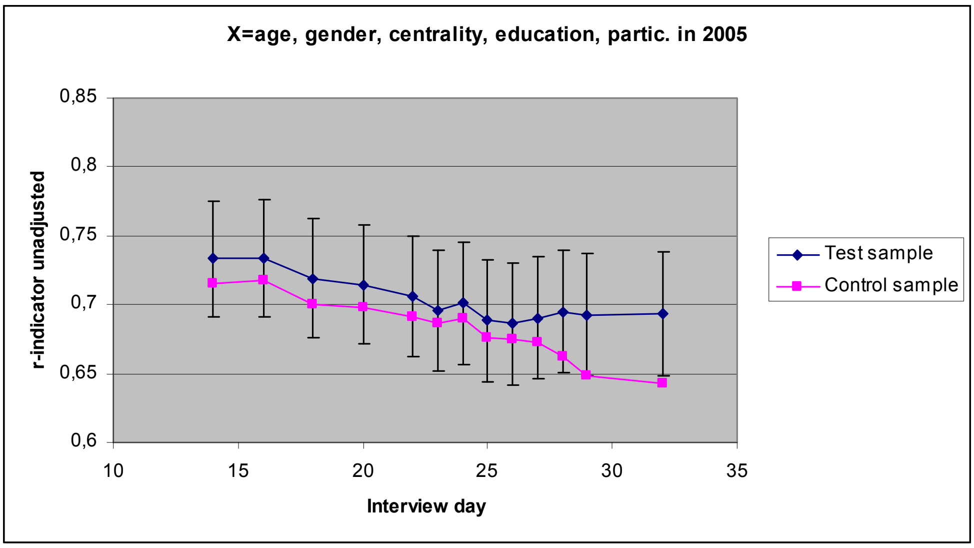

Figure 6 R-indicators for Test sample and Control sample using the five explanatory variables age, gender, centrality, education and status in 2006. Confidence interval for the test sample indicator.

3.4. Discussion Statistics Norway

The R-indicator is a useful tool for evaluating responsive survey design and data collection process.
It has the advantage over the earlier used system that it is a standardised tool with a quality checked software. Further, it can easily introduce more variables, and gain efficiency, without getting overly complex for the user. The interpretation and use of R-indicators requires some training of field staff. In an organisation it is well known that a new management system needs to be learned before it can be productive. Although in our pilot the responsive survey design turned out to not have a great effect, the R-indicator was able to show that an intervention had taken place. We feel that R-indicators can be used as a monitoring device, especially for deciding what kind of respondents to put in the parking lot (it is harder to invent smart strategies to persuade refusers).

4. Introduction Pilot Statistics Netherlands

Statistics Norway used R-indicators and partial R-indicators to decide upon a responsive design during fieldwork. Statistics Netherlands (CBS) used them to design a differential strategy prior to fieldwork. Based on knowledge of response behaviour of similar sample units in the past, probability of contact and cooperation was manipulated by differentiating the timing, spacing and number of contact attempts, and the interviewer assigned to the unit. Aim of the pilot was to augment representativeness of sample realisation, against maximally equal, but ideally less, costs and with minimally equal, but ideally higher, response rates.

4.1. Method

As a vehicle for the pilot the monthly Survey of Consumer Confidence (SCC) was used. This is a CATI survey, conducted among 1500 households of whom a listed telephone number can be found. Questions are asked of any person in the household core (head of household or partner). The length of the questionnaire is about eight minutes. Questions are asked related to sentiments about the household's economic situation and expenditure. Fieldwork is conducted in the first ten workdays of each month.

Because the SCC is conducted monthly, a wealth of information is available about contact and cooperation characteristics of former sample units. This accumulated knowledge was used to determine fieldwork strategy prior to the start of the fieldwork.

The fieldwork of the pilot was conducted during the months of October, November and December 2009. It was conducted alongside the regular SCC, during the same 10 day fieldwork period, with a similar sampling method, a similar sample size and, as far as possible, the same interviewers. The SCC served as control for the response and representativeness measures.

In order to achieve the aim of better representativeness with lower costs, a mixed mode design was chosen for the pilot, in which a mail and/or web first round was followed by a CATI follow-up of nonrespondents. Mail and web questionnaires not only cost less to administer than CATI questionnaires, they can also reach respondents that are otherwise hard to contact and/or to convince to cooperate. Calculation of consumer confidence occurs on data collected within the first ten days of each month. As it is not feasible to conduct a mixed mode design with CATI follow-up within ten days, the design of the pilot was adapted. Figure 7 illustrates the design of the pilot.

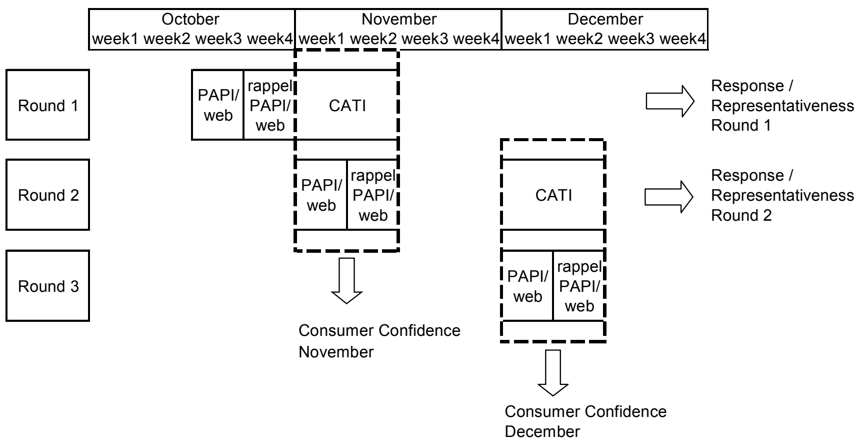

Figure 7. Design of the pilot.

The first mail/web round was conducted during the last fortnight of October[5]. One week after sending the advance letter, a reminder was sent. Ten days later, the CATI follow-up of nonresponse started, which was conducted in the first two weeks of November. Three days before the first of November, advance letters were sent for the second round of fieldwork, again to be followed by a reminder one week later. CATI follow-up of non-respondents of round 2 started on the first of December. As in November, three days prior to the first of December, advance letters and questionnaires were sent to the third sample. This sample received the advance letter and one reminder, but no CATI follow-up. As is shown in figure 1, Consumer Confidence is calculated across two different samples: mail/web response of month T and the CATI follow-up response of month T-1. Response rates and representativeness of the response, are however calculated within one sample (i.e., within each round).

Fieldwork strategy of the pilot was determined based on what could be learned of the response propensities of sample units in two existing datasets. The SCC 2004 (available at www.r-indicator.eu) was used to estimate contact and cooperation propensities for the telephone survey. The dataset of the SCC 2004 contains cooperation and contact information of about 18.000 sample units, as well as auxiliary information, made available from CBS registries. The CBS Safety Monitor 2007 was used to estimate cooperation propensities for the web/mail survey.

4.1.1 Linked data

The samples of both SCC and the pilot were linked to the Social Statistical Database of Statistics Netherlands. This database consists of administrative information on persons, households, jobs, benefits and pensions.

The variables used for the analysis are displayed in Table 6. There is geographical, demographic and socio-economic information on different levels. The lowest level in the registries is the person. In this analysis, however, the level is the household. All person variables are therefore aggregated to a household level, based on information about the household core (head of household and partner). Because of this aggregation, the variables ethnic group and gender have a category to indicate a mixture of the categories on the personal level (e.g., mixed native-foreign). The next level comprises information at the postal code level.

Table 6. Linked data to the Survey of Consumer Confidence

Variable	*Categories*
Household level	
Ethnic Group	Native, Morrocan, Turkish, Suriname / Netherlands Antilles, other non-western, other western, mixed and unknown. For the present analyses aggregated to native, foreign, mixed and unknown
Gender	All male, all female, mixed, unknown
Average age of household core	15-30;31-44;45-65; over 65, unknown
Type of Household	Single, partners without children, partners with children, single parents, unknown
Postal code area level	
Degree of urbanization	Very strong, strong, moderate, low, not urban, unknown
percentage non-western non-natives	Very high, high, average, low, very low, unknown
average monthly income	Quartiles

[5] Because of time constraint issues, and because the data of this first round were not used to calculate consumer confidence, it was decided to do the first round in the last to weeks of October, in stead of the first two weeks. Analyses showed that this had no response implications.

Each variable has a category ‘information not available’. This has to do with linking sample units to registries. As registries are never entirely up to date, people moving, building or demolishing dwellings, and unregistered people may lead to unavailable information both at the level of the individual or household, or the level of the postal code. Rather than treating these absent data as missing values, they are incorporated as meaningful values.

4.1.2 Over- and under-represented groups

Partial R-indicators were calculated to determine which groups are over- or under-represented within SCC 2004. Groups with a high, medium, or low contact propensity and groups with a high, medium, or low cooperation propensity were identified. This propensity was then projected upon the new samples for the pilot.

A simple sum score was used to determine the expected contact and cooperation propensity in the samples of the pilot and control group. For example, the partial R-indicators showed that elderly households, households with low incomes, households of non-Dutch origin, households living in a neighbourhood with a high percentage of people of non-Dutch origin, and single persons were less likely to participate than other households. The more of these elements present in a single household, the lower the chance of cooperation. I.e., an elderly household with a low income would have a lower cooperation propensity than an elderly household with a high income. A similar exercise was done for chance of contact, where it was shown that young households, living alone or in a partnership without children, households living in highly urban areas, households of non-Dutch origin and households living in neighbourhoods with a high percentage of non-Dutch, have a low contact propensity. Again, the propensity is lower, the more elements present. Based on these analyses, each sample unit was classified as having a high, medium or low contact propensity and having a high, medium or low cooperation propensity. Results of the first month showed that the medium cooperation group should be split in two. In the second month, therefore, four groups were differentiated.

The propensity analysis for SCC 2004 was repeated for a Statistics Netherlands’ survey with a mixed mode design (the Safety monitor), to investigate how high, medium or low response propensity in a CATI survey related to response behaviour in a web / mail first round. For the Safety monitor, people were invited to participate in a web survey. They could however request to receive a mail questionnaire.

It was shown that cooperation propensity, as calculated for the CATI SCC data was highly predictive of web response as well. Web response of the people predicted to be relatively ‘easy’, i.e., having a high cooperation probability, was 31,3% in the first wave web round, whereas the ‘hard’ group had a response of 4,8%. However, the group with a low web response, had a relatively high mail response. Mail response in the group we defined as ‘easy’ on cooperation was 6,4%, but 13,5% in the group with the lowest cooperation propensity.

These findings led to the conclusion that both a mail and a web version of the pilot questionnaire were necessary in order to gain cooperation in the hardest group.

4.1.3 Differential fieldwork strategy

Web/mail wave.

On the basis of the predicted web and mail response of the three cooperation groups, the following design was decided upon for the first web/mail wave:

- households with a high chance of cooperation would receive an invitation for the web survey
- households with a medium chance of cooperation would receive an invitation for the web survey and a mail questionnaire. Either could be filled in.

- households with a low chance of cooperation received only a mail questionnaire. This simplified the advance letter to a great extent, and it was expected that that would be beneficial to response.

All households received one reminder. The reminder mentioned that an interviewer would call, if the questionnaire was not received within shortly. No new mail questionnaire was sent along with the reminder.

Telephone wave.

In the second wave, the nonresponse was followed up by CATI. In this wave it was attempted to

1. stimulate chance of contact for sample units with a low contact propensity
2. dampen the number of contact attempts for units with a high contact propensity
3. stimulate cooperation for sample units with a low cooperation propensity, and
4. dampen cooperation for sample units with a high cooperation propensity.

For different groups, different approach strategies were defined in the CATI management system, by means of the definition of different time slices. The CBS CATI management system is a Blaise application. Defining time slices enables the CATI management system to allocate telephone numbers according to criteria that can be different for different time slices. By defining multiple time slices per day, an address can be called more than once a day. Defining different time slices for groups of addresses makes differential fieldwork strategy possible.

One time slice was defined for elderly Dutch households (65 years and older). This group has a high contact propensity, but a low cooperation propensity. To make interviewer capacity available for groups who needed a higher number of contact attempts, the CATI fieldwork for this group was postponed to the second week of fieldwork period. The households were called primarily during daytime. One evening only was reserved for hitherto uncalled numbers in this group: the last night of the fieldwork period. The definition of this time slice not only freed valuable capacity for evening calls, but was also cost effective, as daytime shifts are remunerated 40% less than evening shifts. In the second month of the pilot, the definition of this time slice was slightly adapted however, because the dampening effect was too strong. In the second month, the fieldwork for this group started in the first week of the fieldwork period, and numbers were called on two weekday evenings each week. In the last week a further adaptation was made, to make numbers of this group available on two additional evenings.

The second time slice consisted of single households, households of non-Dutch origin, households in highly urban areas and households consisting of young people (30 years or under). The time slice was to be called in every shift (morning, afternoon and evening), every day of the fieldwork period.

A third time slice consisted of people of 31 to 45 years of age, not belonging to the second time slice. This group was to be called during the evening for the first two contact attempts. Subsequent attempts could be made during the day also. The last time slice was the miscellaneous ‘other’ group. They received the default treatment that the control group, the regular SCC also received.

Although the definition of time slices determines when numbers can be called, whether they are actually called is dependent on the available interviewer capacity in a shift. To assure that if limited capacity was available, numbers of households with the lowest contact propensity would be called with preference, these numbers were prioritized in each day batch by using an algorithm that used the predicted contact probability.

Definition of time slices and prioritizing numbers in a day batch, were measures taken to influence contact probability. In order to influence cooperation probability, the assignment of numbers to specific interviewers was manipulated. Based on their SCC work in 2008 and the first half of 2009, interviewers were classified in three categories, according to the cooperation rates achieved. A top quartile of the best interviewers (mean cooperation rate in 2008-2009: 82,1%), a middle group of the second and third quartile (cooperation rate 74%) and a third group in the lowest quartile (65,6%). The best interviewers called households with the lowest cooperation propensity. The interviewers with the lowest response rated called households with the highest probability of cooperation. The group in between called the middle group. On top of that, if appointments were made for a certain date or time, the appointment would be followed up by an interviewer of the lowest quartile. On the other hand, if a 'soft' appointment was made '*call me back some other time'*, this would be followed up by an interviewer in the best quartile. In the second month of the pilot, the middle group was split in two, to be able to make a finer distinction in the households with a medium cooperation propensity. The assignment of groups of addresses to groups of interviewers was handled by the CATI management system. To prevent planning problems, interviewers of a 'better' quartile would always be allowed to call numbers meant for a 'lower' quartile. In practice, this possibility was seldom used, however. See the Blaise CATI guide (2004) for details of how definition of time slices and allocation of interviewers to addresses may be attained.

4.1.4 Fieldwork in the control group

The regular SCC is a one mode - telephone only- survey. No information is available beforehand of the characteristics of the households. In practice, this means that all households have an equal probability to be selected in the day batch, although households with whom appointments are made are prioritized. 80% of the fieldwork is performed during evening shifts. During daytime shifts, an interviewer is present to call appointments made for daytime, and s/he may use spare time to work other numbers. Supervisors determine daily whether the work advances satisfactory and whether it would make sense to call an address one or more additional times. The basis for this decision is overall response rate. As in the experimental group, an advance letter is sent some days prior to commencing fieldwork. In neither pilot nor SCC incentives were given or promised, and no refusal conversion was attempted.

4.2. Results

4.2.1 Response

Table 7 shows response results for the regular Survey of Consumer Confidence (the control group) and the pilot. Despite the slight changes in the design in the second month, results were highly comparable and are collapsed.

Table 7. Response results of the SCC and the pilot.

	SCC		*Pilot*	
Results	*N*	*Percent*	*N*	*Percent*
Ineligible	225	7,5	144	4,8
Non-contact	196	6,5	183	6,1
Not present during fieldwork period	73	2,4	62	2,1
Not able (ill, dementia)	115	3,8	122	4,1
Language problems	40	1,3	26	0,9
Refusal	467	15,6	548	18,3
Response	1884	62,8	1915	63,8
Response WEB-PAPI			1081	36,0
Response CATI			834	27,8

In both pilot months, the number of response cases was higher in the experimental group. Because of the substantial number of ineligible cases in the SCC, the response rate RR01 (AAPOR, 2006), i.e., the response of eligible cases, was slightly higher in the SCC, however. Ineligible cases in this kind of CATI research consist mostly of disconnected telephone numbers. Disconnected numbers are correlated with a predicted low chance of noncontact and non-cooperation however. It will be shown that sending a mail questionnaire to high-risk addresses contributed substantially to the response of these households, and to a better representative response.

4.2.2 Representativeness

The response results show comparable response rates for pilot and control group. Table 8 shows, by means of R-indicators, the representativeness of this response, as well as that of each steps in the fieldwork process: the representativeness of the eligible part of the sample, of those contacted, of those able, and of those cooperating. The R-indicator ranges from 0 (no representativeness) to 1 (complete representativeness). As can be seen in the table, the R-indicator of each subsequent step is higher in the pilot than in the control group, with the exception of ‘being able to cooperate’. Only for the R-indicator of response do confidence intervals not overlap, however ($p < .05$).

Table 8. R-indicators and 95% confidence interval for eligible, contacted, able, cooperating and responding cases in the SCC and the pilot.

	SCC		Pilot	
	R	CI	R	CI
Eligible	0,84	(0,809 - 0,870)	0,88	(0,851 - 0,910)
Contacted	0,83	(0,796 - 0,862)	0,87	(0,836 - 0,900)
Able	0,86	(0,830 - 0,886)	0,85	(0,827 - 0,882)
Cooperation	0,87	(0,837 - 0,901)	0,89	(0,857 - 0,916)
Response	0,77	(0,738 - 0,804)	0,85	(0,816 - 0,877)*

* *p<.05*

Analysis of the partial R-indicators shows how the experimental manipulations affected sample composition.

Table 9 shows the results of the analysis of unconditional (or bivariate) partial indicators for each step in the fieldwork process. The SCC starts with determining eligibility, whereas the pilot starts with the response on the web/mail round. For each auxiliary variable, the italic value is the composite contribution to representativeness; the other values describe the positive or negative contribution of the categories of the variable. In this section, the term ‘better representativeness’ is used if a value is closer to zero. However, the absence as yet of confidence intervals for partial indicators makes the extent of this being ‘better’ unknown.

Table 9. Unconditional partial indicators for SCC and pilot.

	SCC					Pilot					
	eligible	contact	able	coop-eration	response	response web/mail	eligible	contact	able	coop-eration	response
Age	*59*[1]	*52*	*38*	*35*	*58*	*62*	*43*	*33*	*45*	*21*	*36*
< 30	-26	-41	8	17	-25	-10	-18	-12	1	14	-10
30-44	-17	-13	15	14	1	-34	-11	-23	20	12	0
45-64	13	8	14	7	29	3	9	15	18	-8	22
65>	26	22	-29	-26	-10	43	20	10	-35	-7	-10
no information available	-42	-17	-10	-2	-43	-28	-30	-8	-6	3	-25
Gender	*142*	*108*	*101*	*41*	*209*	*118*	*105*	*71*	*91*	*58*	*134*
Male(s)	-18	-37	-9	-1	-43	13	-5	-30	3	25	-3
Mixed	21	27	27	7	54	2	15	20	20	-9	30
Female(s)	-2	-8	-30	-13	-38	0	-1	-5	-29	-4	-28
no information available	-51	-27	-19	5	-53	-28	-37	-9	-8	0	-34
Household composition	*51*	*49*	*35*	*18*	*88*	*53*	*40*	*38*	*37*	*29*	*52*
Single	1	-28	-23	0	-37	4	0	-27	-31	10	-33
Partners, with children	15	20	13	2	32	37	18	14	12	-16	17
Partners, no children	11	20	16	7	37	-32	3	18	17	-3	25
Single parent	-16	-5	-3	-16	-28	3	-13	-7	0	22	4
no information available	-45	-29	-16	-5	-57	-20	-33	-11	-3	4	-26
Ethnic group	*57*	*33*	*40*	*13*	*71*	*35*	*42*	*15*	*25*	*17*	*43*
Native	14	10	9	-6	15	1	12	5	3	-1	13
Mixed	8	3	15	8	23	21	7	3	14	-8	9
Foreign	-19	-16	-31	7	-38	-2	-16	-11	-19	14	-22
no information available	-51	-27	-19	5	-53	-28	-37	-9	-8	0	-34
Income in quartiles	*51*	*29*	*35*	*14*	*67*	*49*	*37*	*14*	*38*	*14*	*54*
<1600	-1	-4	-19	-5	-21	-10	3	2	-32	-5	-23
1600-1900	5	4	-7	-9	-7	-13	-1	-7	3	-7	-11
1900-2300	4	3	9	1	12	-4	5	5	11	0	14
2300>	10	7	20	9	31	36	9	4	16	11	30
no information available	-49	-28	-18	4	-53	-28	-35	-10	-7	-1	-34
Urban density	*18*	*30*	*18*	*16*	*32*	*28*	*31*	*15*	*11*	*26*	*24*
very strongly urban	-6	-21	-11	9	-19	5	-6	-13	-2	18	0
strongly urban	-11	-6	-8	-1	-16	14	4	2	-5	0	0
medium urban density	6	0	5	7	13	-9	6	4	-1	6	10
low urban density	9	11	9	-9	12	5	5	4	8	-16	-2
No urban density	2	16	5	-6	10	-4	7	4	2	-5	5
no information available	4	5	-2	0	5	-20	-28	-4	-4	1	-21
% non-western foreigners in area	*47*	*38*	*25*	*35*	*60*	*33*	*34*	*18*	*14*	*18*	*25*
Less than 5%	14	17	9	-4	22	14	12	9	-1	-7	8
5-10%	9	3	5	-2	8	6	8	-1	1	1	6
10-20%	-3	-4	1	-16	-15	-9	-1	-15	10	7	0
20% and more	-12	-23	-17	30	-13	-12	-11	-3	-10	14	-6
no information available	-42	-26	-15	-1	-51	-24	-29	-3	0	-4	-22

Table 9 shows for the variable ‘Age’, that better representativeness is attained in the pilot in all columns, with the exception of the column ‘Able’. As to eligibility, the better representativeness of Age in the pilot is attained for all age groups, although the difference is only in the extent to which groups are over- or under-represented. Concerning contact, representation of the under 30 years old and the elderly is better in the pilot, signifying a higher contact rate for the young households, and a lower contact rate for the elderly. The findings for cooperation show that representativeness of the age groups is the same for pilot and control group, with the exception of the elderly, who are better represented in the pilot. In the final column, this is reflected in a better overall representation of Age in the pilot, especially through a better result for the young households and the households of which no information is available. The variable ‘Gender’ show the largest difference between the pilot and the SCC. Both eligibility, contact and ability are more representative in the pilot, but cooperation is not, as a result of an over-representation of households consisting of one or more males. Together with under-representation in contact for this group, this translates in a virtually perfect representativeness of male households in the response, however. Female households are less under-represented, households of unknown gender composition are less under-represented and households of mixed gender are less over-represented in the pilot. Similar findings are found for the variable 'Household composition', in which all categories appear to be better represented in the response of the pilot. The variable 'Ethnic background' shows that ethnic minorities are perfectly represented in the web/mail round. Eligibility, contact and ability are better represented, especially for the households of whom no auxiliary information was available. Cooperation showed no difference, while response showed a better representativeness in the pilot. Native Dutch were equally well represented in the pilot and the SCC, but both ethnic minorities, households of mixed ethnic background and households of which ethnic composition was unknown were better represented in the pilot. The variables 'Income' and 'Urban density were hardly affected by the experimental manipulations. Ethnic composition of the neighbourhood was better represented in all fieldwork steps. The better representation in response was attained for all groups to some extent, but, again, especially for the households of which no information was available.

To illustrate how unconditional R-indicators relate to traditional analyses of (non)response, Table 10 shows bivariate analyses of the relation between auxiliary variables and eligibility, contactability, etc., in the pilot and the regular SCC. The values in Table 10 should be compared to the italic values in Table 9, showing the overall contribution to representativeness of the variable.

Table 10. Bivariate analyses (Cramèr's V) of eligible, contacted, able, cooperating, and responding cases in SCC and Pilot.

	Eligible		*Contacted*		*Able*		*Participation*		*Response*		*Response Web/Mail*
Variable	*SCC*	*Pilot*	*SCC*	*Pilot*	*SCC*	*Pilot*	*SCC*	*Pilot*	*SCC*	*Pilot*	*Pilot*
Age	0.25	0.22	0.18	0.13	0.15	0.17	0.08*	0.05 ns	0.13	0.09	0.13
Gender	0.22	0.19	0.19	0.15	0.16	0.14	0.04 ns	0.06*	0.20	0.11	0.06*
Household composition	0.20	0.19	0.18	0.15	0.12	0.15	0.05 ns	0.07*	0.18	0.11	0.11
Ethnic group	0.21	0.20	0.12	0.05*	0.13	0.10	0.03 ns	0.04 ns	0.15	0.09	0.07
Income	0.19	0.17	0.10	0.05 ns	0.12	0.15	0.04 ns	0.04 ns	0.14	0.11	0.10
Urban density	0.07*	0.14	0.11	0.06 ns	0.06 ns	0.04 ns	0.04 ns	0.06 ns	0.07*	0.05 ns	0.06 ns
Percentage non-western foreigners	0.18	0.16	0.14	0.07*	0.08*	0.05 ns	0.08*	0.04 ns	0.12	0.05 ns	0.07

** p < .05; ns No significant relation; all other values p<.01*

It can be seen that results of the two analyses are highly comparable. Whenever the partial R-indicators show a larger deviation from representativeness, Cramèrs V is larger. The bivariate analysis does not show, however, what the contribution of the respective subgroups to the deviation is, as the partial R-indicators do. Response matrices could in theory fill that gap. However, as is shown in section 4.2.5, response matrices quickly become cumbersome to interpret, especially with a large number of auxiliary variables. For example, bivariate response analysis shows that singles, elderly people and females have a low response rate. The low response rate for singles could be related to the response rate of (single) elderly. The same could be true for the low response of (single) females. These relations will need to be studied in multiple cross tables. The same holds for ethnic minorities and people living in neighbourhoods with a high ethnic minority density and people living in the cheapest houses. They may well be one group. The number of relevant cross tabulations is high, and may never lead to satisfactory interpretations. Differences in the size of the subgroups may further complicate interpretation. Conditional partial R-indicators on the other hand, are weighted for group size and clearly show which of the effects remain when controlled for the other variables.

The results of table 10, as described above, show that in all probability, variables are interconnected. The better representativeness of singles, males and elderly, for example, may very well be a better representativeness of the group of elderly single males. To analyse if this is the case, conditional R-indicators can be calculated, that correct for the other auxiliary variables in the model. Table 11 shows these conditional R-indicators. In contrast to the unconditional indicators, the conditional indicators can no longer be illustrated by the response matrix. Nor can the direction of the deviation from representativeness be gauged.

The conditional R-indicators show that, when corrected for the other variables in the model, the deviation from representativeness is not substantial, neither for the SCC, nor for the pilot. The largest deviation is for males in the SCC, and for the highest incomes in the pilot, but neither are very large. Nevertheless, also in this analysis, the representativeness of the variables in the pilot is slightly better. Gender still is the variable with the largest deviation from representativeness in the SCC, as in the unconditional analysis. The relatively large deviations in household composition and ethnic group, as observed in the unconditional analysis, all but disappear when controlled for the other variables. In the pilot, the largest deviation is observed for Income, caused by the large deviation (over-representation) of the highest incomes in the web/mail round.

Table 11. Conditional partial R- indicators for SCC and pilot.

	SCC					Pilot					
	eligible	contact	able	coop-eration	response	response web/mail	eligible	contact	able	coop-eration	response
Age	*31* (1)	*40*	*26*	*29*	*24*	*37*	*21*	*31*	*29*	*16*	*13*
< 30	32	67	4	17	25	7	16	5	1	9	3
30-44	27	30	12	14	5	51	12	56	21	5	8
45-64	14	20	18	17	21	24	6	18	27	9	6
65>	23	41	31	35	6	52	10	15	36	1	1
no information available	1	1	2	10	2	0	0	1	1	4	1
Gender	*14*	*13*	*17*	*10*	*31*	*15*	*3*	*14*	*11*	*21*	*12*
Male(s)	9	9	8	4	40	15	0	11	8	21	6
Mixed	7	4	9	3	27	1	0	1	0	2	1
Female(s)	5	5	13	4	26	6	0	7	3	21	8
no information available	0	0	0	0	0	0	0	0	0	0	0
Household composition	*9*	*16*	*14*	*24*	*22*	*25*	*8*	*21*	*15*	*22*	*18*
Single	2	10	3	14	6	2	1	17	6	6	14
Partners, with children	1	5	4	7	4	26	2	7	11	11	5
Partners, no children	1	4	5	4	4	25	1	16	2	5	4
Single parent	2	3	2	16	4	2	3	1	3	21	11
no information available	2	3	4	19	31	7	0	2	1	3	1
Ethnic group	*8*	*5*	*19*	*6*	*17*	*15*	*8*	*4*	*18*	*9*	*15*
Native	2	1	11	1	8	6	2	1	7	2	7
Mixed	0	1	2	2	3	17	0	1	2	4	2
Foreign	3	1	24	1	18	1	4	1	24	3	14
no information available	0	0	0	0	0	0	0	0	0	0	0
Income in quartiles	*5*	*5*	*17*	*13*	*21*	*36*	*4*	*10*	*26*	*20*	*28*
<1600	1	1	9	2	13	26	0	1	41	7	24
1600-1900	1	1	3	6	5	21	1	2	6	9	15
1900-2300	0	0	2	1	2	8	0	2	7	2	8
2300>	0	0	14	8	22	76	0	1	11	20	28
no information available	0	0	0	0	0	2	0	4	0	1	2
Urban density	*13*	*13*	*11*	*9*	*14*	*18*	*17*	*7*	*10*	*18*	*14*
very strongly urban	2	1	1	0	0	3	28	3	3	8	8
strongly urban	7	3	3	0	10	8	0	0	2	1	1
medium urban density	2	2	1	4	5	13	0	0	1	4	3
low urban density	3	2	3	3	2	3	0	0	4	15	6
No urban density	2	10	1	0	1	3	0	0	1	1	1
no information available	0	0	3	0	1	3	1	0	0	2	2
% non-western foreigners in	*7*	*9*	*5*	*30*	*15*	*15*	*6*	*11*	*12*	*11*	*5*
Less than 5%	1	1	0	8	2	5	0	2	2	2	1
5-10%	3	1	0	3	1	1	1	0	0	0	0
10-20%	1	1	1	21	7	4	0	6	8	1	0
20% and more	0	4	1	41	3	1	1	2	2	2	1
no information available	1	1	1	13	10	13	0	1	2	7	1

The equivalent of the conditional R-indicators would be a multivariate logistic regression. To illustrate the R-indices in Table 11, Table 12 shows the logistic model for the regression on response of all auxiliary variables for the SCC and the pilot. As representativeness in the RISQ project is defined as 'absence of predictable contribution', non-representative groups would show up in a standard (logistic) regression, as well as in the analysis of the conditional R-indicators. In the multivariate logistic regression on response 'Gender', 'Household' and 'Age' are selected in the model for the SCC, while 'Gender' and 'Income' are selected for the pilot. Again, conclusions of the two analyses are comparable, although not identical. The larger the multivariate deviation from representativeness in the conditional R-indicators, the more the log odds in the logistic regression deviate from 1. However, in the pilot, the largest deviation is shown for Income, followed by Household Composition, which does not show up in the multivariate model. For the SCC, the model shows the three variables with the largest deviation, but the order is slightly different.

Table 13 shows the final multivariate model. The model fit for both is poor, with a low pseudo R^2 of 6.3% for the SCC, and an even lower 2.4% for the pilot. The lower fit for the pilot is another indication of better representativeness.

Table 12. Logistic model for the response propensity in the SCC and pilot.

SCC					Pilot				
Variable	*Category*	*β*	*Exp(B)*		*Variable*	*Category*	*β*	*Exp(B)*	
Gender	Male(s)	-0,91	0,40	***	Gender	Male(s)	-0,17	0,85	ns
(reference = mixed)	Female(s)	-0,73	0,48	***	(reference = mixed)	Female(s)	-0,36	0,7	***
	No info	-0,92	0,40	***		No info	-0,41	0,67	ns
Household	Partners, no children	0,02	1,02	ns	Income	2nd	0,14	1,15	ns
(reference = partners with children)	Single	0,16	1,18	ns	(reference = lowest quartile)	3rd	0,23	1,26	ns
	Single parent	-0,13	0,88	ns		Highest quartile	0,44	1,55	ns
	No info	0,72	0,49	**		No info	0,55	1,74	ns
Age	30-45 y.o.a.	-0,22	0,80	*					
(Reference = 45-65 years of age)	Less than 30 y.o.a.	-0,59	0,56	**					
	Over 65 y.o.a.	-0,21	0,81	*					
	No info	-							

****** p < .001; ** p < .01; * < .05; ns not significant*

Table 13. Multivariate analysis of responding versus un-responding cases in the SCC and pilot.

Variable		*Wald χ^2*		*Variable*		*Wald χ^2*
Gender	111.23	31.61	34.31	Gender	36.32	15.20
Household		14.49	13.53	Income		16.34
Age			11.80			
pseudo R^2	0.051	0.057	0.063	pseudo R^2	0.016	0.024
χ^2	114.58	128.99	140.77	χ^2	36.25	52.62
df	3	7	10	df	3	7

4.2.3 Results of the experimental manipulations

The pilot consisted of three manipulations: adding a mode, manipulation of contact chance in the CATI part, and manipulation of chance of participation, again in CATI. This paragraph describes the effect of these measures on the subsequent distribution of response. Response, cooperation and contact rates are used to illustrate the effect of the manipulations on representativeness, as described in the previous paragraphs.

<u>Adding a mode</u>

The effects of the experimental manipulation on eligibility have to do with the role of the web/mail first round of data collection in the pilot. To be able to influence the distribution of response, web/mail results will have to be different from the CATI distribution. Table 14 shows that they are. The table shows response rates for web and CATI results for the auxiliary variables. In the right hand of the table, these response rates are indexed within each variable, to facilitate interpretation. Especially the category ‘no information available’ benefited from the mixed mode design. But also males and single parents were better represented as a result of the added mode. Adding - especially - mail as a mode resulted in a very high cooperation of elderly in the first round (reflected in the categories ‘over 65’, ‘male(s)’, ‘female(s)’, and ‘single’). This high cooperation did not eventually result in over representation however, due to the dampening measures that were taken in the subsequent CATI round (calling mostly during the day, and starting the fieldwork in the second week).

Table 14. Response rates of SCC and web/mail and CATI rounds in pilot.

		Response SCC	Web/mail pilot	CATI-pilot	Response pilot	Response SCC	Web/mail pilot	CATI-pilot	Response pilot
Age	45-65jr	67,4	36,5	30,9	67,4	100*	100	100	100
	30-45	62,9	28,9	35,3	64,3	93	79	114	95
	Less than 30	52,4	30,8	27,9	58,7	78	84	90	87
	Over 65	61,0	44,0	18,0	62,0	90	120	58	92
	No information available	34,0	23,9	25,2	49,1	50	65	81	73
Gender	Mixed	69,6	36,3	31,4	67,8	100	100	100	100
	Male(s)	50,4	39,6	23,2	62,9	72	109	74	93
	Female(s)	54,7	36,0	22,1	58,1	79	99	70	86
	No information available	34,0	23,9	25,2	49,1	49	66	80	72
Household	Partners with children	69,6	30,2	38,0	68,3	100	100	100	100
	Partners without children	68,3	42,5	24,3	66,8	98	141	64	98
	Single	55,5	36,9	20,3	57,2	80	122	53	84
	Single parent	50,3	37,2	28,5	65,7	72	123	75	96
	No information available	40,1	28,4	25,7	54,1	58	94	68	79
Ethnic group	Natives	64,5	36,1	29,2	65,3	100	100	100	100
	Mixed	70,6	43,1	23,7	66,8	109	119	81	102
	Foreign	49,6	35,4	20,4	55,8	77	98	70	85
	No information available	34,0	23,9	25,2	49,1	53	66	86	75
Income in quartiles	Less than 1600	58,1	33,6	24,8	58,4	84	78	93	84
	1600-1900	61,5	33,3	28,2	61,5	89	77	106	88
	1900 - 2300	65,1	35,3	31,4	66,7	95	82	118	96
	More than 2300	68,9	43,0	26,6	69,6	100	100	100	100
	No information available	34,6	24,1	25,3	49,4	50	56	95	71
% non-western foreigners in area	< 5%	65,6	37,9	26,9	64,8	100	100	100	100
	5-10%	65,1	37,9	27,6	65,5	99	100	103	101
	10-20%	58,2	33,2	30,7	63,9	89	88	114	99
	>20%	58,7	32,0	29,9	61,9	89	85	111	96
	No information available	39,0	27,3	28,6	55,8	59	72	106	86
Urban density	Highly urban	58,0	37,3	26,5	63,8	89	106	89	98
	Urban	59,7	39,2	24,5	63,7	91	112	82	98
	Medium	65,7	34,0	32,3	66,2	100	97	108	102
	Low urban density	65,4	37,2	26,2	63,4	100	106	88	98
	Not urban	65,5	35,1	29,8	64,9	100	100	100	100
	No information available	65,2	28,1	27,6	55,6	100	80	93	86

**Reference category = 100.*

Manipulating chance of contact

The higher R-index for the contact phase shows that the manipulations of contactability were successful in attaining a somewhat better representative contacted sample. Table 15 illustrates these findings with the contact rates for SCC, compared to total contact rates for the pilot and the contact rates for the CATI part of the pilot separately. The table shows that, with the exception of the group with estimated high contact propensity, contact rates were generally somewhat higher in the pilot than in the SCC, albeit not significantly so according to Chi2 analysis. LSD posthoc tests showed that in the SCC the 14 percentage points difference between the groups with estimated low and high risk of noncontact was significant. Also, the group with the highest contact propensity had indeed a significantly higher contact rate than the other groups. In the pilot however, the difference between the groups with estimated lowest and highest contact propensity was not significant, neither in the CATI part of the pilot, nor in the complete pilot data. Also, no difference in contact rate existed between the three groups with the lowest contact propensity.

Table 15. Contact rate per propensity category for SCC, the CATI part of the pilot and pilot total.

Noncontact propensity	Pilot - CATI	Pilot total	SCC	χ^2
Low noncontact propensity	91.4a	94.5a	95.1a	ns
Low medium noncontact propensity	86.3b	91.6b	88.5bc	ns
High medium noncontactvpropensity	80.8b	86.6c	85.3cd	ns
High noncontact propensity	84.4ab	87.8abc	81.1d	ns

χ^2 analysis of the difference between contact rate of pilot-total versus SCC within propensity groups.
Values with shared subscripts within one column do not differ significantly according to LSD post-hoc tests.

Analyses of the realisation of the calling strategy showed that:

1. addresses with a high risk of noncontact were attempted significantly sooner, low risk addresses significantly later than similar addresses in the pilot than in the SCC. For example, 62% of high risk addresses was called in the first or second day of the fieldwork period for the pilot, against 36% for the SCC. On the other hand, 12% of low risk addresses had their first attempt in the last three days of the fieldwork period, against 0% in the SCC.
2. a comparable number of calls was made in the pilot and SCC for low risk addresses; a significantly higher number of calls was made to higher risk addresses in the pilot however (4.2, 5.1, and 5.0 versus 2.8, 3.0 and 3.5 for lower medium to high risk addresses in the pilot and SCC respectively).
3. The distribution of calls was different for the pilot and SCC, with significantly more daytime calls, less evening calls, and more calls per day.

<u>Manipulating chance of participation</u>

Chance of participation was manipulated by having the best interviewers call addresses with the highest chance of non-participation and vice versa. Analysis of the fieldwork verified that the fieldwork strategy was applied as planned and that the mean level of interviewer capacity was comparable in the pilot and the SCC. Although the number of participating sample units was slightly higher in the pilot, the participation rate was somewhat lower ($\chi^2 = 4{,}23$, $p < .05$). The R-indicator for participation, table 9, showed that there was hardly any difference in the distribution of participation for the pilot and control group. Table 16 below illustrates this finding with the participation rates per propensity group for the pilot, its CATI part, and the SCC. Having the best interviewers call the hardest cases did not bring about the expected rise in cooperation, but having the lesser interviewers call the easy cases brought about a significant decline in cooperation, resulting in an even distribution of participation across the four propensity groups.

Table 16. Participation rate by propensity group for the CATI part of the pilot, the pilot total and SCC.

Participation propensity	Pilot-CATI	Pilot-total	SCC	χ^2
High participation propensity	65.6a	76.9a	82.5a	p < .01
Higher medium propensity	54.2b	76.2a	79.8ab	ns
Lower medium propensity	66.0a	81.1a	76.4b	ns
Low participation propensity	61.2ab	78.4a	78.2ab	ns

χ^2 analysis of the difference between participation rate of pilot-total versus SCC within propensity groups.
Values with shared subscripts within one column do not differ significantly according to LSD post-hoc tests.

Some light is shed on the issue of why the best interviewers were not able to secure a higher cooperation rate by studying table 17. This table shows response results for pilot and SCC by participation propensity.

Table 17. Response by participation propensity for pilot and SCC.

	Pilot				SCC			
	high	*high medium*	*low medium*	*low*	*high*	*high medium*	*low medium*	*low*
Ineligible	2,4	3,8	9,4	4,8	4,4	6,7	9,3	8,1
Noncontact	3,1	4,7	10,8	7,8	3,4	4,5	11,2	9,4
Not able (ill, not present)	2,7	2,5	6,9	12,3	3,5	4,5	7,3	12,1
Language problems	0,0	0,5	1,1	2,4	0,2	0,6	1,0	4,3
Refusal	21,2	20,8	13,3	15,7	15,5	16,9	16,8	14,4
Response	70,6	66,7	57,0	56,9	73,1	66,9	54,4	51,7
Response Web/mail	30,6	43,8	35,1	36,7				
Response CATI	40,0	22,8	21,9	20,2				
N	995	744	639	619	1.100	674	493	630
Cooperation rate	74,7	72,8	71,4	65,1	79,2	75,3	68,4	62,7
Cooperation rate 2	76,9	76,2	81,1	78,4	82,5	79,8	76,4	78,2

As can be seen, the cooperation rate is higher, as predicted, as the estimated participation propensity is higher, in the pilot as well as the control group. Prediction of participation propensity was based on the calculation of cooperation according to COOP2 (AAPOR 2006), as cooperation of contacted eligible sample units, However, as table 17 shows, prediction of participation appears to be heavily correlated with both ability to participate, as with the existence of language problems. The second cooperation rate that is calculated in the table shows cooperation of eligible, contacted, ànd able sample units (COOP3). With this calculation, the differences between groups that were estimated to be of low, medium or high risk of non-participation, all but disappear. If the only difference between groups in the level of cooperation is related to the ability to participate, a different intervention is needed, for example using translated questionnaires and bi-lingual interviewers.

4.2.4 Costs

One of the aims of this pilot was to augment data quality while ideally diminishing costs. To compare the costs of the pilot to that of the control group, only the actual costs of observation and subsequent data processing (for the mail questionnaire) are considered. Two measures were taken that would introduce a substantial amount of costs saving: the use of a web round, and the larger share of day-time interviewing, as interviewers at Statistics Netherlands receive a 40% higher remuneration for evening work. Whether the use of a mail questionnaire would diminish costs, compared to a CATI version, was not clear beforehand, because of uncertainty concerning the number of respondents that would choose mail over web, and the subsequent amount of data handling necessary.

Without counting the allowance for evening work, the pilot turned out to be 18% cheaper than the regular SCC. Counting the 40% rise for evening work, the difference was 22%.

4.2.5 Comparing partial R-indicators to subgroup response rates.

In paragraph 4.2.2 is described how partial R-indicators and especially conditional partial R-indicators give better insight in the specific subgroups that need attention than response rates do. Bivariate response rate matrices may give misleading information, while cross tabulations may quickly become hard to read and interpret. In this section an analysis of response rates is compared with an analysis of partial R-indicators. The data used are of the 2005 Survey of Consumer Confidence, the same data used to calculate response propensities for the pilot. Table 18 shows bivariate response rates. It is shown that elderly people, people living in neighbourhoods with a high number of non-western non-Dutch inhabitants, people of non-Dutch origin, first generation non-Dutch ethnic groups, singles, people in inner cities, people living in the cheapest quartile of houses, people renting their home, males and females have a low response rate. In all cases, response rates of people or households of which no information is known is very low.

***Table 18**. Response rates of subgroups in the Survey of Consumer Confidence, 2005.*

		response	N
Age	<30 y	66,0%	1.134
	30-44	69,7%	4.914
	45-64	69,7%	6.900
	65 and over	58,2%	4.960
Percentage non-western non-Dutch in neighbourhood	< 5%	68,9%	61
	5-10 %	66,6%	11.893
	10-20 %	67,0%	2.020
	> 20%	62,4%	1.763
Worth of house	lowest quartile	61,4%	4.471
	2nd quartile	66,8%	4.453
	3rd quartile	68,4%	4.462
	highest quartile	68,8%	4.452
	no information	47,1%	70
Ethnic group	Native	67,3%	14.324
	Non-Dutch	56,6%	987
	Mixed	68,3%	2.210
	no information	41,6%	387
Generation	Native	67,3%	14.324
	First generation	51,0%	608
	second generation	63,3%	354
	Mixed 1st and 2nd generation	68,7%	2.235
	no information	41,6%	387
Household composition	Single	56,7%	4.906
	Partners without	68,6%	6.056
	Partners with children	73,2%	5.564
	Single parent	66,9%	756
	Other / no information	57,2%	628
Urban density	Very urban (>=2500 adresses/km2)	62,0%	2.844
	Urban (1500 tot 2500 adresses/km2)	66,8%	3.961
	Medium (1000 tot 1500 adresses/km2)	67,5%	3.455
	Low (500 tot 1000 adresses/km2)	67,5%	3.692
	Not urban (<500 adresses/km2)	66,9%	3.913
	no information	41,9%	43
Ownership house	Ownership	69,9%	10.983
	Rental home	60,7%	6.827
	no information	50,0%	98
Gender	Male(s)	56,8%	2.058
	Mixed	70,7%	11.419
	Female(s)	57,9%	3.826

The low response rate for singles could in part be explained by the low response rate of the elderly. The same could be true for the low response of males and females (who are predominantly single). These relations are depicted in table 19. Table 19 shows that not all singles are underrepresented; the response of 44-65 year old female singles is above average. The table also shows that not all elderly are under represented: married elderly couples have a relatively high response rate.

Table 19. Response of male, female and mixed households by age and household type.

			Response	N
Single	< 30 y	Male	50,9%	110
		female	53,3%	137
	30-45	Male	54,7%	516
		female	60,8%	291
	45-65	Male	59,9%	568
		female	66,1%	726
	>65	Male	57,3%	558
		female	52,3%	1.922
Unmarried partners without children	< 30 y	Male(s)	-*	
		mixed	71,6%	348
		female (s)	-	
	30-45	Male(s)	38,5%	26
		mixed	64,2%	321
		female (s)	-	
	45-65	Male(s)	68,0%	25
		mixed	66,3%	181
		female (s)	-	
	>65	Male(s)	-	
		mixed	62,9%	62
		female (s)	-	
Married without children	< 30 y	Male(s)	-	
		mixed	75,5%	151
		female (s)	-	
	30-45	Male(s)	-	
		mixed	70,4%	304
		female (s)	-	
	45-65	Male(s)	-	
		mixed	71,4%	2.477
		female (s)	-	
	>65	Male(s)	-	
		mixed	65,1%	2.002
		female (s)	-	
Unmarried partners with children	< 30 y	Male(s)	-	
		mixed	77,5%	40
		female (s)	-	
	30-45	Male(s)	-	
		mixed	73,0%	344
		female (s)	-	
	45-65	Male(s)	-	
		mixed	70,8%	96
		female (s)	-	
	>65	Male(s)	-	
		mixed	-	
		female (s)	-	
Married with children	< 30 y	Male(s)	-	
		mixed	73,1%	156
		female (s)	-	
	30-45	Male(s)	-	
		mixed	75,3%	2.495
		female (s)	-	
	45-65	Male(s)	-	
		mixed	72,1%	2.143
		female (s)	-	
	>65	Male(s)	-	
		mixed	61,7%	120
		female (s)	-	
Single parent	< 30 y	Male	-	
		female	70,4%	27
	30-45	Male	78,6%	28
		female	67,5%	209
	45-65	Male	60,7%	84
		female	71,4%	238
	>65	Male	-	
		female	59,1%	115

** less than 25 observations*

Table 20 shows the relationship between age and urban density, to study whether the under-representation of elderly is less in less urban areas. Contrary to expectation, response rates are lower for all elderly, but in rural areas, the difference is larger than in urban areas. The information in both tables needs to be combined to study response of elderly married couples by urban density. Table 20 also shows that, although the response of the young households is generally not worrisome, that of the young households in urban regions is. By focussing on bivariate response rates, this information may be missed. Crossing this finding with household information is necessary to show whether all young households are under represented in urban regions. This is shown in table 21. It appears from this table that not all young households are under represented: the response of young urban couples is quite high. On the other hand, the number of cases is too small to merit a separate approach, and is too small to have a large influence on data quality.

Table 20. Response of age groups by urban region.

				Response	N
Urban density	Very urban (>=2500 adresses/km2)	age	<30 y	60,2%	279
			30-44	64,3%	800
			45-64	65,2%	881
			65 and over	57,4%	884
	Urban (1500 tot 2500 adresses/km2)	age	<30 y	68,4%	253
			30-44	69,6%	1.004
			45-64	69,0%	1.481
			65 and over	61,5%	1.223
	Medium (1000 tot 1500 adresses/km2)	age	<30 y	65,5%	194
			30-44	71,4%	949
			45-64	71,6%	1.363
			65 and over	58,2%	949
	Low (500 tot 1000 adresses/km2)	age	<30 y	65,9%	205
			30-44	71,7%	1.049
			45-64	69,8%	1.491
			65 and over	59,6%	947
	Not urban (<500 adresses/km2)	age	<30 y	72,3%	202
			30-44	70,4%	1.092
			45-64	71,1%	1.676
			65 and over	54,3%	943

These tables only scratch the surface of the number of cross-tabulations possible. Furthermore, a differential fieldwork strategy dictates that similar sets are made for noncontacts and non-cooperation. By using (partial) R-indicators, however, the interrelation between variables can easily be studied, without the difficulty in interpretation that multiple response matrices pose.

Tables 21 to 23 show this, as well as the zooming-in function of the R-indicator. Table 21 starts with a 'wide angle' view of the data, the overall R-indicator for the 2005 SCC response. Table 22 zooms in on the variables influencing representativeness, both unconditionally and conditionally. Table 23, finally, fully zooms in at the within variable level. The column with unconditional partial R-indicators within this table is comparable with response rate analyses. The column with conditional indicators, however, shows the relative impact of the variable, conditional upon the other variables in the response model. The conditional partial R-indicator isolates that part of the departure of representativeness that can be attributed solely to a specific variable. Because the number of cases within a given variable is accounted for in the calculation, if the value of a partial R-indicator is large, the impact of the variable is large, even if the group considered is small.

Table 21. R-indicator, Standard Error and Confidence Interval of the response of the Survey of Consumer Confidence.

R-indicator	SE	Confidence interval
0.83	0.007	0.825-0.839

The R-indicator of 0.83 is well within the range Statistics Netherlands Surveys usually show (from .80 to .85).

Table 22. Unconditional and conditional Partial R-indicators of the response of the Survey of Consumer Confidence.

Variable	Unconditional	Conditional
Household type	0.0654	0.0109
Age	0.0497	0.0281
% not-western in neighbourhood	0.0136	0.0056
Worth of house	0.0315	0.0046
Ethnicity	0.0452	0.0332
Urban density	0.0244	0.0100
Gender	0.0613	0.0078

The unconditional partial R-indictors in table 22 show that Household type is the variable with the largest bivariate impact on response, closely followed by Gender. The conditional partial R-indicators however, show that the impact of both variables is far less when controlling for the other variables in the model. Conditionally, Ethnic origin appears to be the variable with the largest impact on representativeness, followed by Age. Zooming in allows to study these variables in more detail. Table 23 shows that, apart from the addresses where no information is available, the impact of western foreigners is large in this dataset. Although this is a small group, the impact is noticeable. Also, being Dutch has a relatively large impact. The indicators for age suggest that the largest influence is exerted by the elderly and the group of 45 to 65 year olds.

Table 23. Unconditional and conditional Partial R-indicators of the within variable categories in the response model of the Survey of Consumer Confidence.

Variable	Unconditional	Conditional
Household		
Single	-0.050	0.005
partners no children	0.013	0.006
partners with children	0.037	0.006
single parent	0.009	0.002
other / no information	-0.012	0.005
Age		
< 30	0.0006	0.0054
30-45	0.0183	0.0098
45-65	0.0202	0.0147
>65	-0.0415	0.0212
% not-western in neighbourhood		
<5 %	0.0022	0.0021
5-10%	0.0026	0.0021
10-20%	0.0027	0.0043
> 20%	-0.0127	0.0019
No information available	0.0022	0.0012
Worth of houses in neighbourhood		
Cheapest quartile	-0.0239	0.0034
2nd quartile	0.0021	0.0020
3rd quartile	0.0108	0.0009
Most expensive quartile	0.0126	0.0019
No information available	-0.0117	0.0011
Ethnic origin		
Dutch	0.0086	0.0126
Non-western foreign	-0.0131	0.0044
Western foreign	-0.0215	0.0167
Mixed	0.0086	0.0054
No information available	-0.0355	0.0249
Urban density		
Very urban (>=2500 addresses/km2)	-0.0162	0.0050
Urban (1500 - 2500 addresses/km2)	0.0030	0.0068
Medium urban (1000 - 1500 addresses/km2)	0.0051	0.0022
Little (500 - 1000 addresses/km2)	0.0053	0.0013
Not (<500 addresses/km2)	0.0022	0.0045
No information available	-0.0163	0.0015
Gender		
Male(s)	-0.0318	0.0060
Mixed	0.0357	0.0028
Female(s)	-0.0384	0.0041

The conditional partial R-Indicators show that the low response rate in a number of groups is in fact caused by other factors. For example, being single appears not to be a factor in it self, nor is urban density. However, as will be described in the discussion, more work is needed in order to be able to correctly interpret the value of these partial indicators. Presently, confidence intervals are being developed, that enable determining when a contribution to (non) representativeness is actually no longer significant.

4.3. Discussion Statistics Netherlands

The pilot at Statistics Netherlands aimed at obtaining a better representative response, against lower costs. In the control group, a CATI Survey of Consumer Confidence (SCC) was held, using a uniform fieldwork strategy. In the experimental group, a mixed mode differential fieldwork strategy was deployed. Previous rounds of the SCC were used to calculate partial R-indicators, identifying groups that are over- or under-represented in contact and / or cooperation. A fieldwork strategy was designed to either stimulate or discourage contact and / or cooperation.

Results show that the differential fieldwork strategy was successful in maintaining the level of response, while significantly augmenting representativeness and at the same time substantially reducing costs. The R-indicators showed that representativeness was especially augmented as a result of more representative eligible- and contacted cases in the pilot. The manipulation of cooperation had less impact.

Analysis of unconditional and conditional partial R-indicators made possible detailed consideration of the influence of the manipulations on different groups. For example, it was shown that the experimental manipulations had a large effect on the groups of whom no information was available. Also, young households were better represented as a result of measures to stimulate contact. The net representation in the response of the elderly households did not differ in the pilot and the control group, but the manipulations to dampen contact in this group, while at the same time stimulating cooperation, were visible in the partial R-indicators.

5. Overall discussion

This paper describes two exercises using R-indicators and partial R-indicators in the field, in an endeavour to ameliorate representativeness of the response, either by responsive design, or by pre-designed differential fieldwork strategies. Although both pilots encountered some difficulties, these difficulties had more to do with the design of the fieldwork than with the use and interpretation of the R-indicators. While the overall R-indicator showed to which extent we were successful in our manipulations, the partial R-indicators proved very useful in either determining where to put effort, or deducing why interventions did not go as planned. Although the unconditional R-indicators are more illustrative of which groups are under- or over- represented, and to what extent representation deviates, the conditional R-indicators need to be studied with great care at all times, in order to prevent unforeseen effects in groups that you did not plan to manipulate.

In this paper, partial R-indicators were calculated alongside subgroup response rates. The partial R-indicators allowed for zooming in on relevant subgroups in a way that subgroup response rates could not. The main differences between subgroup response rates and R-indicators are (Schouten en al., 2010a):

1. "Partial R-indicators are linked to R-indicators, i.e. they represent the contribution of variables to the lack of representativeness, while subgroup response rates are linked to response rates. In other words they are conceptually different, although they have response propensities as basic ingredients.
2. Partial R-indicators are available at the variable-level.
3. Partial indicators are computed both unconditionally and conditionally.
4. Partial R-indicators are weighted differences between subgroup response rates and overall response rates. The weight is proportional to the size in the population. Subgroup response rates are not weighted."

The second difference makes R-indicators and partial R-indicators fit for zooming in and out.

The third and fourth difference make partial R-indicators effective in searching for groups that help improving representativeness of survey response.

The researchers involved in both pilots were all enthusiastic about the potential of these indicators. Of course, more research is needed. First, it is crucial to get an understanding of acceptable levels or the indicators. Second, within the RISQ project, additional work is currently undertaken to calculate confidence intervals. Partial R-indicators have a precision that depends on the sample size. In order to increase the relevance of the indicators, they need to be supported by error margins. Third, in the Dutch contribution to this paper some attention is paid to the comparison of partial R-indicators with traditional analytical measures that express the relation between response (or contact, or cooperation) and (auxiliary) variables. This kind of evaluation is essential in assessing and understanding the usefulness of the R-indicators.

Apart from the two pilots, considerable effort was spent in RISQ to investigate the possibilities of R-indicators and partial R-indicators for monitoring data collection. Corresponding papers were developed under work package 6 of the project and are available at the website www.risq-project.eu. Schouten et al., (2010a) summarises all practical applications of the indicators developed within RISQ.

At the RISQ website, code in R and SAS as well as a graphical tool can be downloaded, and these enable computation of R-indicators and partial indicators at both the variable and category level. Two additional releases of the R and SAS programmes are planned. In November 2010 standard errors and confidence intervals will be added. In May 2011 population-based R-indicators will be added. Population-based R-indicators enable the computation of indicators when, in addition to the net sample, auxiliary variables are only known through aggregated population tables.

References

AAPOR (2006), "Standard definitions: Final dispositions of case codes and outcome rates for surveys", 4th ed., Discussion Paper Kansas: AAPOR, The American Association for Public Opinion Research.

Couper, M.P. (1998), "Measuring Survey Quality in a CASIC Environment." *Proceedings of the Survey Research Methods Section, American Statistical Association*, 1998, pp. 41-49.

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

Hapuarachchi, P., March, M. & Wronski, A. (1997), "Using Statistical Methods Applicable to Autocorrelated Processes to Analyze Survey Process Quality Data," in Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N & Trewin, D., eds, *Survey Measurement and Process Quality*, Wiley, pp. 589–600.

Mohl, C. & Laflamme, F. (2007), "Research and responsive design options for survey data collection at Statistics Canada", Proceedings of ASA Joint Statistical Meeting, Section 293, July 29 – August 2, Salt Lake City, USA.

Scheuren, F. (2001), "Macro and Micro Paradata for Survey Assessment," paper presented at the United Nations Work Session on Statistical Metadata, Washington, DC.

Schouten, B. & Bethlehem, J. (2009), "Representativeness indicators for measuring and enhancing the composition of survey response", RISQ deliverable, www.risq-project.eu.

Schouten, B., Cobben, F. & Bethlehem, J. (2009), "Indicators for the representativeness of survey response", *Survey Methodology*, 35 (1), 101 – 113.

Schouten, B., Luiten, A., Loosveldt, G., Beullens, K. & Kleven, Ø. (2010a), "Monitoring and changing data collection through R-indicators and partial R-indicators", RISQ deliverable, www.risq-project.eu.

Schouten, B. & Shlomo, N. (2010b), "Indicators for Representative Response", Proceedings of the Q2010 conference, Helsinki, May 2010.

Shlomo, N., Skinner, C., Schouten, B., Bethlehem, J. & Zhang, L.C. (2009a), "Statistical properties of representativity indicators", RISQ deliverable, www.risq-project.eu .

Shlomo, N., Skinner, C., Schouten, B., Carolina, T. & Morren, M. (2009b), "Partial Indicators for Representative Response", RISQ deliverable, available at http://www.risq-project.eu/papers

Sundvoll, A., Lagerstrøm, B. O., Kleven, Ø. & Löfgren, T. (2008), "Management Tools for Enhancing the Composition of Survey Response", Paper presented at The European Conference on Quality in Official Statistics. Rome, July 2008.

Thomsen, I., Kleven, Ø., Wang, J.H. & Zhang, L.-C. (2006), "Coping with decreasing response rates in Statistics Norway. Recommended practice for reducing the effect of nonresponse", Reports 2006/29, Statistics Norway.

Wagner, J. (2008), "*Adaptive survey design to reduce nonresponse bias*", PhD thesis, University of Michigan, USA.

Westat (2004). *Blaise CATI guide*, Rockville, MD.