RISQ

Representativity Indicators for Survey Quality

**Indicators and data collection control
Work plan and preliminary findings –
Pilot Statistics Netherlands**

Work package 7, Deliverable 8.2

Annemieke Luiten & Willem Wetzels
Centraal Bureau voor de Statistiek, the Netherlands

Januaty 8th, 2010

SEVENTH FRAMEWORK
PROGRAMME

Indicators and data collection control;

work plan and preliminary findings

1. Introduction

In the recent survey literature a lot of attention has been devoted to the level of effort in a survey and the so-called continuum-of-resistance model, see e.g. Fitzgerald and Fuller (1982) and Lin and Schaffer (1995). In the continuum-of-resistance model, households and enterprises are thought to behave along two dimensions, ease-of-contact and ease-of-participation. Attached to those dimensions are individual contact and response probabilities, and, when combined, overall individual response probabilities which form the basis of our perception of representativeness. The contact and co-operation probabilities are unknown but can be modelled using available auxiliary information and the fieldwork paradata.

Associated with the continuum-of-resistance model is the level of effort invested by the survey organisation. The more effort the survey researcher invests in contacting households and converting reluctant respondents, the higher the response rate. Seemingly, the level of effort invested has increased during the past decades in many countries in order to maintain acceptable response rates. As a consequence, the costs of surveys per sampled unit have also increased. It is, therefore, of great importance how the additional efforts are allocated, i.e. the efficiency of these efforts becomes of a growing importance. In fact, the level of effort represents survey costs and can be balanced to response rates, see e.g. Kalsbeek et al. (1994).

In the literature the implications of increased efforts are often debated. Apart from an increased risk of measurement errors, it is also questioned whether the efforts lead to more quality and a more representative set of respondents. Clearly, if difficult-to-contact or difficult-to-convert individuals are different from other individuals, then a focus on easy-to-contact and easy-to-convert units will increase the contrast between respondents and non-respondents. Consequently, the response rate may have increased but the non-response error may not have changed. For instance, a follow-up using telephone interviewing may help raising response rates, but can only be applied to households with a listed phone number. Hence, a single-minded increase of the level of effort may not help improving the quality of the response.

One may, therefore, differentiate the level of effort between households and enterprises to get a balanced, representative composition of the response, see Groves and Heeringa (2005), Biemer and Link (2006) and Van der Grijn, Schouten and Cobben (2006). In determining the level of effort needed for a certain household or enterprise, Representativity Indicators may serve as useful tools. The differentiation of the level of effort may be directed at increasing the response rate while maintaining or even enhancing the representativeness of the response.
Groves and Heeringa (2005) propose so-called responsive designs, which are designs that are dynamic with respect to the composition of the response, i.e. they aim at controlling the response to a survey during the data collection. One may also decide to differentiate beforehand, so that different sampled units are assigned different fieldwork protocols based on historic fieldwork paradata.

In work package 7 we investigate the Representativity Indicators and partial Representativity Indicators as tools to facilitate differentiated fieldwork strategies before and during the data collection phase. We investigate how to differentiate strategies and how the indicators can play a role in the corresponding choices. We examine how differentiated data collection strategies rely on models for the estimation of response probabilities. In order to validate and test the differential fieldwork strategies, two pilot studies were set up one at Statistics Netherlands (CBS) and one at Statistics

Norway (SSB). This paper describes the setup of the field study CBS is presently conducting. A separate paper describes the field study executed by SSB.

2. Method

As a vehicle for the pilot the monthly Survey of Consumer Confidence 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.[1] Length of the questionnaire is about eight minutes, in which questions are asked related to sentiments about the household's economic situation and expenditure. (add appendix with the questions?). Fieldwork is conducted in the first ten workdays of each month. The response of the SCC is about 66%, the percentage of noncontact about 10% and the refusal rate is 23%.

Because the SCC is conducted monthly, a wealth of information is available about contact and cooperation characteristics of former sample units. While SSB concentrates the pilot on responsive design, e.g., adapting fieldwork strategy according to results during fieldwork, CBS used the accumulated knowledge of prior surveys to determine fieldwork strategy prior to the start of the fieldwork.

The fieldwork of the CBS pilot was conducted during the months of October, November and December 2009. It was conducted alongside the regular SCC, during the same fieldwork period of ten days, with a similar sampling method, a similar sample size of 1500 households, and, as far as possible, the same interviewers. The regular SCC served as control for the response and representativeness measures.

One of the aims of the pilot is to augment representativeness of sample realisation, against minimally equal, but ideally less, costs and with minimally equal, but ideally higher, response. In order to achieve this aim, a mixed mode design was chosen, in which a PAPI and/or web first round was followed by a CATI follow-up of nonrespondents. PAPI or web questionnaires not only cost less to administer than CATI questionnaires, but have the added advantage to be able to reach respondents that are otherwise hard to contact and/or to convince to cooperate. Because of the PAPI/web phase, the design of the SCC had to be adapted. Calculation of consumer confidence occurs on data collected within the first ten days of each month. As it is hardly feasible to conduct a mixed mode design with CATI follow-up within ten days, the design was adapted so that consumer confidence was calculated on the PAPI/web response of month T, and the CATI response of month T-1. Figure 1 illustrates the design of the pilot.

---

[1] Head of household or partner

Figure 1. Design of the CBS pilot.
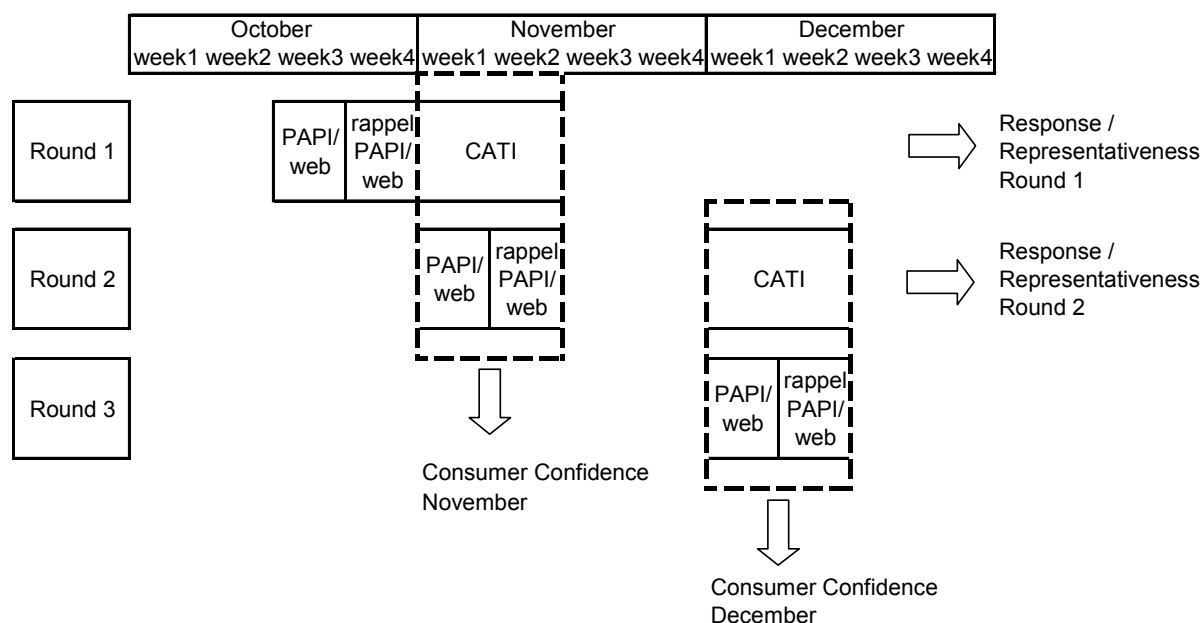
The first PAPI/web round was conducted during the last fortnight of October[2]. One week after sending the advance letter, a reminder was sent. Again one week later, the CATI follow-up of nonresponse started, which was conducted in the first two weeks of November. Three days before the first of November, advance letters were sent for the second round of fieldwork, again to be followed by a reminder one week later. The web server was closed after ten workdays, on the last day of the CATI fieldwork of the control group. Some mail questionnaires were returned after the closing of the official fieldwork period, however. CATI follow-up of non-respondents 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. Response rates and representativeness of the response, are however calculated within one sample (i.e., within each round).

Fieldwork strategy 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 determine contact and cooperation propensities in the telephone survey. The CBS Safety Monitor 2007 was used to determine cooperation propensities in a web/PAPI 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.

2.1 Linked data

The samples of both SCC and the experimental SCC (ESCC) were linked to the Social Statistical Database of Statistics Netherlands. This database samples of both SCC and the experimental SCC (ESCC) were linked to the Social Statistical Database of Statistics Netherlands. This database consists of administrative information on persons, households, jobs, benefits and pensions. It covers the entire Dutch population. The population register, denoted in Dutch by GBA or Gemeentelijke Basis Administratie, serves as the backbone for the SSB. The GBA contains mostly demographic information for all the persons that are or were registered in one of the municipalities since 1995.

---

[2] 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. Contrasting results with round two will determine whether this decision has response implications.

The variables used for the analysis of (E)SCC are displayed in table 2.1. There is geographical, demographic and socio-economic information on different levels. The lowest level that is used in this analysis is the household level. All person variables are therefore aggregated to the household level, based on information about the household core. Because of this aggregation, the variables ethnic group and gender and household type have an additional category, to indicate a mixture of the categories on the personal level. The next level comprises information at the postal code level. The postal code consists of four digits and two letters. Together with the house number, the postal code is unique for every address[3].

Table 2.1 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 or mixed
ethnic generation	native, first generation, second generation, mixed
Gender	all male, all female, mixed
Average age of household core	15-30;31-44;45-65;>65
Type of Household	Single, partners without children, partners with children, single parents, mixed or unknown
Postal code area level	
Degree of urbanization	very strong, strong, moderate, low, not
Percentage non-natives	very high, high, average, low, very low
percentage non-western non-natives	very high, high, average, low, very low
average house value	quartiles

2.2 Over and under represented groups

Work package 6 describes how partial R-indicators can be calculated to determine which groups are over or under represented in sample realisation. This technique was used to determine groups within SCC2004 with a high or low contact propensity and groups with a high or low cooperation propensity. Using a simple sum score, this propensity was then projected upon the new samples for the pilot. For example, the partial R-indicators showed that elderly people, people with low incomes, people of non-Dutch origin, especially the first generation, and people living in a neighbourhood with a high percentage of people of non-Dutch origin, and single persons were less likely to participate than other people. Not shown by the partial R-indicator, but shown by subsequent research was that the more of these elements present in a single household, the lower the chance cooperation. I.e., an elderly person with a low income would have a lower cooperation propensity than an elderly person with a high income). A similar exercise was done for chance of contact, where it was shown that young persons, living alone or in a partnership without children, people living in highly urban areas, people of non-Dutch origin and people living in neighbourhoods with a high percentage of non-Dutch, have a high risk of being non-contacted. Again, the chance is higher, the more elements present. Based on these analyses, each sample unit was classified as being of high, medium of low risk of being non-contacted and as having a high, medium of low risk of not-cooperating. Results of the first round showed that the medium cooperation group should be split in two. In the second round, therefore, four groups were differentiated.

The propensity analysis for SCC 2004 was repeated for the Safety monitor, to determine response behaviour on the web / PAPI first wave of this survey of the 'hard' and 'easy' groups identified above.

---

[3] This passage is adapted from Cobben (2009).

For the Safety monitor, people received an advance letter with a login code for the web survey. They could also request to receive a PAPI questionnaire. The advance letter was followed by two reminders. 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 we predicted to be relatively 'easy', i.e., having a high cooperation probability, had a response of 31,3% on 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 PAPI response. PAPI response in the group with the highest cooperation propensity was 6,4%, while 13,5% in the group with the lowest cooperation propensity.

These findings led to the conclusion, that if we were to save costs by using a mixed mode, both a PAPI and a web version of the pilot questionnaire were necessary, in order to gain cooperation in the hardest group.

2.3 Differential fieldwork strategy

On the basis of the web and PAPI response of the three cooperation groups, the following design was decided upon for the first 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 PAPI questionnaire. Either could be filled in.
- households with a low chance of cooperation received only a PAPI 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 questionnaire was send along with the reminder.

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 time slices were defined in the CATI management system. 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. In each time slice, a number will receive one call attempt[4].

One time slice was defined for elderly Dutch people (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 started in the second week of fieldwork period only. These households were called primarily during daytime. Only the last night of the of the fieldwork period, non-contacted numbers in this time slices were made available for interviewing. The definition of this time slice not only freed valuable capacity for evening calls, but was again cost effective, as daytime shifts are remunerated 40% less than evening shifts. In the second month or the pilot, the definition of this time slice was slightly adapted however, because the dampening effect was too strong. 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.

---

[4] Within one call attempt, multiple dial attempts can be made, for example if a number is engaged.

A third time slice consisted of people of 31 to 45 years of age, not belonging to the second time slice. This groups 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 highest chance of noncontact would be called, 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 highest risk of non-cooperation. 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 fieldwork period, 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 handbook (reference) for details of how definition of time slices and allocation of interviewers to addresses may be attained.

2.4 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 daybatch, 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 both ESCC en SCC no incentives were given nor promised, and no refusal conversion was attempted.

3. Preliminary results

At the time of writing, the second fieldwork period is just finished, and some preliminary findings can be given of overall response results and response in the groups that are over and under represented in the SCC, according to the 2004 data. No results of the multivariate representativeness are as yet available, nor of the comparison of representativeness of the control and experimental group.

3.1 Response

Table 3.1 shows response results for the two rounds for the experimental and control groups.

Table 3.1 Response results of the experimental and regular SCC

| | Round 1 | | | | Round 2 | | | |
| | ESCC | | SCC | | ESCC | | SCC | |
Results	N	Percent	N	Percent	N	Percent	N	Percent
Ineligible	46	3,1	104	6,9	98	6,5	121	8,1
Eligible	1.454	100,0	1.396	100,0	1.402	100,0	1.379	100,0
Non-contact	124	8,5	119	8,5	59	4,2	77	5,6
Not present during fieldwork period	31	2,1	35	2,5	31	2,2	38	2,8
Not able (ill, dementia)	65	4,5	61	4,4	57	4,1	54	3,9
Language problems	16	1,1	21	1,5	10	0,7	19	1,4
Refusal	256	17,6	229	16,4	289	20,6	238	17,3
Response	962	66,2	931	66,7	956	68,2	953	69,1
Response WEB-PAPI	529	36,4			552	39,4		
Response CATI	430	29,6			404	28,8		

Some (sketchy) observations:
- Response results in the experimental and control groups are comparable when RR01 is calculated (AAPOR, 2004). RR01 is response of eligible cases. When response from sample is calculated, response in the experimental groups was slightly higher: 64,1% in the first round for the experiment, versus 62,1% in the control group. In the second round 63.7 versus 63.5. In both rounds, the number of responses is higher in the experimental group than in the control group. This is a result of the fact that in both rounds the number of ineligible cases is (substantially) higher in the control group. Ineligible cases in this kind of CATI research consist mostly of disconnected telephone numbers. As the results in table 3.2 and 3.3 confirm, the percentage of disconnected numbers correlates with auxiliary variables and with the predicted chance of noncontact and non-cooperation. Sending a PAPI questionnaire to high-risk addresses contributed substantially to the response of these households, and probably to a better representative response.
- The response of the first (web-PAPI) waves was high, especially in view of the short fieldwork period, and the fact that one reminder only was sent. 55% and 58% of responses were attained in the cheaper mode in round 1 and 2 respectively.
- The noncontact rate in the experimental group in round 1 is quite high, as a result of the decision to call a sub-sample in the second week of the fieldwork period only. In the second round, this group was called in the first week as well, with a lower non-contact rate as a result.
- The only further substantial difference between the experimental and control groups, is that the number of refusals is higher in the experimental groups. Whether this is the result of the mixed mode design, or the experimental manipulation of interviewer-assignment, remains to be analysed.

Table 3.2 shows response results by risk of non-participation. To counter risk of non-cooperation, interviewer assignment was manipulated. Some findings:

- The predicted cooperation propensity was quite accurate: the response rate in each propensity group diminishes with estimated cooperation propensity: the lower the estimated risk of non-participation, the higher the response.
- Especially noteworthy was the high response rate in the first wave of both rounds, of the groups that were considered to have a high risk of non-participation. In both rounds, the lowest response in the first wave was attained by the group with the highest cooperation propensity. This group had WEB as the only proposed mode. The highest response in wave one in both rounds was achieved by the group with a relatively high cooperation propensity

(medium 1). This group could do the web questionnaire, but also received a PAPI questionnaire. In this group, as well as in the group 'medium2', PAPI was by far the more popular mode: over 91% of responses were PAPI.
- The lower response rate in wave 1 in the group with the lowest risk is compensated by a higher response in the second CATI wave.
- The calculated cooperation rates shed more light on the response mechanisms in the four groups. The first cooperation rate is CR01 (AAPOR, 2004), that is the number of responses of eligible households that are contacted. The second cooperation rate calculates the inverse of the actual refusals, that is the number of responses of the eligible, contacted and able households. The contrast between these two figures shows that non-cooperation in the groups with higher risk of cooperation is not a result of higher refusals in this group, but of a higher incidence of not being able to participate, because of illness or language problems. Whether the low refusal rate in this groups is a result of the experimental manipulation of assigning the best interviewers, remains to be analysed.
- The middle group (medium 1 and medium 2) have a different response behaviour, and should therefore be considered to be different groups. In the second round, medium 1 and medium 2 were differentiated, with a better result for the latter group as a consequence.

Table 3.2 Response results by risk of non-participation

	Risk of non-participation							
	Round 1				Round 2			
	low	*medium 1*	*medium 2*	*high*	*low*	*medium 1*	*medium 2*	*high*
Ineligible	2,1	2,2	5,8	3,1	2,7	5,4	12,5	6,7
Eligible	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Non-contact	4,4	7,2	15,5	10,7	1,9	2,5	8,6	5,4
Not present during fieldwork period	2,6	1,7	2,9	1,3	1,1	1,4	3,0	4,3
Not able (ill, dementia)	0,8	1,9	6,5	11,7	1,1	2,3	6,3	8,9
Language problems	0,0	0,3	1,4	3,6	0,0	0,8	1,0	1,4
Refusal	21,2	18,2	14,4	14,9	22,3	25,2	14,9	18,2
Response	71,1	70,8	59,2	57,9	73,6	67,7	66,2	61,8
Response WEB-PAPI	30,3	44,1	36,8	36,9	32,4	47,0	40,4	40,4
Response CATI	40,8	26,7	22,4	21,0	41,2	20,7	25,8	21,4
N	516	371	294	319	479	373	345	300
Cooperation rate	74,3	76,3	70,1	64,9	75,1	69,5	72,5	65,3
Cooperation rate 2	77,0	79,6	80,4	79,6	76,7	72,9	81,6	77,2

Table 3.3. shows results by risk of non-contact. To counter risk of non-contact, time slices were manipulated to influence time, number and priority of calls. Some results:

- the result that is most related to risk of non-contact, more than the actual non-contact rate, is the number of ineligible cases. Up to 28% of the high risk group appeared to be disconnected.
- Because of this result, the calculation of response rates is highly influenced by eligibility. Table 3.4 shows the same results, with sample as denominator (in stead of sample minus ineligible).
- Despite the experimental manipulation, the non-contact rate in the groups with a high risk of non-contact is still substantially higher than in the low risk groups. Comparison with the control group and analysis of the history files with calling history should show whether the manipulation had any effect at all.

Table 3.3 Response results by risk of non-contact, eligible cases as denominator

	Risk of non-contact					
	Round 1			Round 2		
	low	*medium*	*high*	*low*	*medium*	*high*
Ineligible	2,1	1,7	12,5	4,2	10,5	27,9
Eligible	100,0	100,0	100,0	100,0	100,0	100,0
non-contact	7,7	9,5	13,5	3,4	6,3	12,5
not present during fieldwork period	2,0	3,5	0,8	1,9	2,3	6,3
not able (ill, dementia)	4,5	4,3	4,5	4,4	3,9	0,0
language problems	0,3	0,0	9,8	0,4	2,3	2,5
refusal	19,5	13,9	10,5	21,2	16,4	18,8
Response	66,0	68,8	60,9	68,7	68,8	60,0
Response WEB-PAPI	36,9	36,4	32,3	38,9	49,2	30,0
Response CATI	29,1	32,5	28,6	29,7	19,5	30,0
N	1.113	235	152	1.246	143	111

Table 3.4 Response results by risk of non-contact, sample as denominator

	Risk of non-contact					
	Round 1			Round 2		
	low	*medium*	*high*	*low*	*medium*	*high*
Ineligible	2,1	1,7	12,5	4,2	10,5	27,9
non-contact	7,5	9,4	11,8	3,3	5,6	9,0
not present during fieldwork period	2,0	3,4	0,7	1,8	2,1	4,5
not able (ill, dementia)	4,4	4,3	3,9	4,2	3,5	0,0
language problems	0,3	0,0	8,6	0,4	2,1	1,8
refusal	19,1	13,6	9,2	20,3	14,7	13,5
Response	64,6	67,7	53,3	65,8	61,5	43,2
Response WEB-PAPI	36,1	35,7	28,3	37,3	44,1	21,6
Response CATI	28,5	31,9	25,0	28,5	17,5	21,6
N	1.113	235	152	1.246	143	111

These results give a preliminary and sketchy impression of the first result. A lot remains to be done. In the following months, representativeness analyses will be done, in which it is shown how the measures described here, have affected sample composition. Also, the costs of fieldwork and control groups will have to be calculated.