

The representativeness of the LISS panel

Marike Knoef

CentERdata and Tilburg University

and

Klaas de Vos

CentERdata

July 2008

1. Introduction

When interpreting the results of analysis of survey data an important question concerns the representativeness¹ of the data. When one is interested in political opinions, consumption patterns, social-psychological constructs or many other topics, it makes a lot of difference if the survey upon which results are to be based may be considered to be representative of the population in question or not. In particular when conclusions are expressed in quantities (averages, percentages, amounts etc) their interest increases considerably when representativeness can be taken for granted. On the other hand, when the survey is not representative, it is difficult to draw convincing quantitative conclusions, and methods to correct for non-representativeness are fraught with complications. One popular solution is to use weighted estimates. However, the calculation of relevant weights is by no means trivial (see e.g. Groves et al 2004, Biemer and Christ, 2008).

Internet surveys use by far the fastest and cheapest data collection method available but when – as is quite common – respondents can register themselves voluntarily, self-selection implies that the survey is unlikely to be representative for the population in question (see e.g. Bethlehem, 2006). Even if the target population is meant to consist of internet users only, the most motivated groups are bound to be overrepresented. When the target population includes non internet users, any claim of representativity can hardly be taken seriously.

Recently, a new internet panel survey, the LISS panel, was started. One of the distinguishing features of this panel is that it explicitly aims to be representative for the (Dutch) population, a.o. by providing a free PC and internet access to households who were not yet on line. In this paper we examine to what extent the panel succeeds in being representative for the population surveyed. We compare the composition of the panel with that of the population of households in the Netherlands, subdivided according to characteristics such as age, gender, region, household size and composition, and education.

Once they have agreed to become members of the LISS panel, the respondents are asked to fill out a questionnaire once per month. The topics of these questionnaires vary. Among them are core questionnaires on income, health, housing and employment which

¹ There is no universally accepted definition of representativeness of surveys (see Stoop, 2006). In this paper we will implicitly define a survey to be representative when all members of the target population have an equal probability to be included in the data to be analyzed, so that estimates based on the data are not biased as a result of systematic under- or overrepresentation of various population groups.

will be repeated annually, but also questionnaires on various topics suggested by scientists with an interest in free, representative data. Unfortunately, panel members sometimes fail to fill out the monthly questionnaire. Therefore, next to results for the panel as a whole we also present a first analysis of the representativeness of the response to individual questionnaires. Obviously, even if the panel is representative, if there are systematic differences between various household groups in the response to individual questionnaires, one may still question whether research results may be considered to be representative for the population. A multivariate regression of the response variable shows whether the response probability differs significantly between persons with different characteristics and provides indications which groups should be the focus of special attention in the data collection process.

2. Sample framework

This paragraph describes the sampling method employed in the LISS panel. The reference population for the LISS panel is the Dutch speaking, non-institutional population permanently residing in the Netherlands. The intention is that the LISS panel consists of about 5.000 households. To compile these households a random, nation-wide, (gross) sample of 10.600 addresses has been drawn from the population registers of the municipalities. In the gross sample unoccupied houses, non-existing addresses, language problems, and a 60% to 70% readiness to participate in the panel have been taken into account. The sample unit is the address, as it is the intention to build a household panel including all members of the households living at a given address. At first, the households were informed with an announcement letter in combination with a brochure and a 10 euro note. Next, respondents were contacted by an interviewer in a mixed mode design and were asked to participate in the LISS panel. Those households for which a telephone number was known were contacted by telephone. The remaining households were visited by an interviewer and thus contacted face-to-face. If a household could not be reached by telephone after several attempts the address was transferred to face-to-face recruiting. With regard to the face-to-face recruiting also several attempts have been made. In case a household refuses to participate, an extensive refusal conversion procedure has taken place, tailored to the type of refusal. More information on the recruitment of the panel can be found on the CentERdata website soon.

In case a household takes part in the LISS panel, one of the household members answers a general questionnaire about some basic characteristics of the household and their household members (the "household box"). As from then all household members are in the full LISS panel. Furthermore, all household members of age 16 and older

indicate whether they want to participate in the monthly questionnaires or not. All persons of age 16 and older who have indicated that they want to answer the monthly questionnaires are participating members of the LISS panel.

For our analysis we have to notice that a (full) household is included in the LISS panel when at least one of the household members aged 18 or older is willing to participate. This could imply that households with more persons have a higher probability to be in the LISS panel than a single person household.²

3. Comparison LISS and national statistics

As explained above, the target population consists of all Dutch private households where the adult persons master the Dutch language sufficiently. In this section we investigate whether the characteristics of the LISS panel match the characteristics of the target population by comparing statistics of the LISS panel with the national statistics. It is not possible to find exact statistics of the precise target population, because in the national statistics the Dutch households with all adult persons not skilful in the Dutch language are included, whereas they are not included in LISS. We will have to take this into account when comparing the national statistics with LISS³. Furthermore, LISS is a household sample with only private households included. The statistics of the Netherlands are sometimes only available for the total population (including persons living in institutions).

Table 1 presents descriptive statistics of LISS and comparisons with the national statistics. For these descriptives (and in the remainder of the paper) we will use the LISS panel as of April 30, 2008. The national statistics are not of this date, they are often from January 1 2007, for some variables they are from January 1 2008, and with regard to education they are from January 1 2006. On April 30 2008 the LISS panel consisted of 4,975 households who have answered the general questionnaire about their household characteristics, containing in total 12,626 people.

² For example, assume that every person has (the same) probability $p \in [0,1]$ to participate. Then, a single person household has a probability of p to participate and a household of two adult persons has a probability of $p^2 + 2p(1 - p)$ to participate in the LISS panel. Obviously, in reality, the willingness to participate of various household members may be correlated.

³ According to SIM 2006 (a survey about the integration of ethnic minorities, "Survey integratie minderheden") about 24% of the Turkish people in the Netherlands have often/always problems with reading the Dutch language. For Moroccans this appears to be 19%, whereas for Surinamese and Antilleans this only holds for 2% (Jaarrapport Integratie 2007, SCP).

From Table 1 we see that the elderly are underrepresented⁴, especially the elderly women. Possibly this has to do with the unfamiliarity of elderly with internet and computers in general. In addition health may play a role.

Table 1. Gender and age groups in LISS, compared with the national statistics

Age	LISS			CBS 2007 (private households)		
	Male %	Female %	Total %	Male %	Female %	Total %
<20	13.7	14.0	27.7	12.5	11.9	24.4
20-39	12.1	13.1	25.1	13.3	13.2	26.6
40-64	19.1	19.5	38.6	17.7	17.4	35.1
65-79	4.0	3.7	7.8	5.0	5.8	10.7
80+	0.5	0.3	0.8	1.1	2.0	3.1
Total	49.4	50.6	100	49.6	50.4	100.0

Table 2 shows the distribution of the number of household members in the households of LISS and in the Dutch population in 2007. Single person households are underrepresented. This may have to do with the sample framework and/or with specific characteristics of single person households. For example, there will be a connection between the underrepresentation of elderly and single person households.

Table 2. Number of household members in LISS, compared with the national statistics

# household members	LISS	CBS 2007 (private households)
	%	%
1	23.7	35.3
2	35.9	32.6
3	13.5	12.5
4	18.9	13.5
>=5	8.0	6.0

Table 3 shows the cross tabulation of age and marital status. For the age group 20-65 married and divorced people are better represented than widowers and never married persons. To some extent this may be related to the previous table, since widowers and never married people will often be single persons.

Table 3. Marital status and age groups in LISS, compared with the national statistics

Age	Married %	Divorced %	Widow(er) %	Never married %	Total %
LISS					
<20	0.0	0.0	0.0	27.7	27.7
20-39	9.9	1.0	0.0	14.2	25.1
40-64	29.2	4.3	0.8	4.2	38.6
65-79	6.1	0.5	0.9	0.2	7.8
80+	0.5	0	0.3	0.0	0.8
Total	45.7	5.8	2.0	46.5	100.0

CBS 2007

⁴ In all tables of this paper we have tested over- and underrepresentation with a chi-square goodness of fit test. In case we mention that a specific group is over- or underrepresented this always means that there is a significant over- or underrepresentation.

CBS 2006 (population)						
Primary education	12.4	4.4	6.5	9.3	13.0	8.9
Secondary education: VMBO	41.3	14.9	18.6	22.0	27.7	24.3
Secondary education: HAVO/VWO/MBO	38.6	46.5	45.7	39.8	35.4	41.4
Tertiary education: HBO	6.4	21.1	17.2	18.3	15.4	15.9
Tertiary education: University	1.3	13.1	12.0	10.6	8.6	9.4
Total	100.0	100.0	100.0	100.0	100.0	100.0

In Table 5 we present the level of urbanization of the participants in LISS and the population of the Netherlands in 2006. The population living in regions with the highest level of urbanization is underrepresented. This may have to do with the population of these areas, but also with the fact that the municipality registers in these areas contain more wrong addresses. A high urbanization level may also be related with a relatively high proportion of single person households, of which we know that they are underrepresented in the LISS panel. Striking is that also the least urbanized regions are underrepresented.

Table 5. Level of urbanization in LISS, compared with the national statistics

	LISS	CBS 2006 (population)
Level of urbanization (number of addresses per km ²)	%	%
>=2.500	12.9	19.2
1.500-2.500	25.7	22.9
1.000-1.500	22.2	18.0
500-1.000	23.0	19.6
<500	16.2	20.4
Total	100.0	100.0

Table 6 shows the presence of people in the various regions of the Netherlands. In the LISS panel relatively fewer people are living in the western part of the Netherlands. In the western part of the Netherlands the age group 0-19 is relatively less overrepresented than in the other parts of the Netherlands. Also, the age group 20-44 is underrepresented in the west, while they are somewhat overrepresented in the other parts of the Netherlands. The underrepresentation of the west may be linked with the urbanization level shown above, as urbanization is highest in the western part of the Netherlands.

Table 6. Region of residence in LISS, compared with the national statistics

Region of residence	North	East	South	West	Total
LISS	%	%	%	%	%
0-19	3.4	6.8	5.9	11.7	27.7
20-44	3.7	7.4	7.7	15.0	33.7
45-64	3.3	6.6	7.7	12.4	30.0
65-79	0.8	1.9	1.8	3.3	7.8
80+	0.1	0.1	0.2	0.4	0.8

Total	11.3	22.7	23.3	42.7	100.0
CBS 2008 (population)					
0-19	2.5	5.4	5.0	11.2	24.0
20-44	3.4	7.1	7.0	16.4	34.0
45-64	2.9	5.7	6.2	12.5	27.3
65-79	1.2	2.3	2.6	4.9	11.0
80+	0.4	0.7	0.8	1.8	3.8
Total	10.4	21.2	21.6	46.7	100.0

4. Filling in the monthly questionnaires

In section 3 we have compared the households in LISS with the national statistics. Every month the members of these households, who are 16 years and older and have indicated that they want to participate in the individual questionnaires, receive one or more questionnaires. In this section we want to investigate the response to two of these individual questionnaires, namely “Religion and ethnicity” (from now on called REE) and “Health” (HEALTH). The response to these individual questionnaires is important. Even if the panel is representative, if the response to the individual questionnaires differs among various household groups, the research results are not representative for the population. In the remainder of this section we will relate the outcome of the REE questionnaire with national statistics, to examine the representativeness of people from different ethnicities in LISS. After that we will investigate the representativeness of the people who answered the REE questionnaire and the HEALTH questionnaire.

In Table 7 we see that especially non-western immigrants and 1st generation immigrants are somewhat underrepresented. Splitting the descriptives for men and women we see that especially the male non-western and first generation immigrants are underrepresented (Table 8). As mentioned before, we have to take into account that LISS only includes households where the adult persons master the Dutch language sufficiently. When looking more specifically to the country of origin it appears that especially Turkish and Moroccan people are underrepresented, while for example people from the Netherlands Antilles are not underrepresented at all (Table 9). This indicates that language probably plays a role. Table 10 shows the presence of immigrants in several age groups. Interesting is that also non-western immigrants in the youngest age groups are underrepresented, although language problems may be assumed to play less of a role for this group. Probably other factors are important as well. Kalwij (2007), for example, found in the first wave of SHARE (Survey of Health, Ageing and Retirement in Europe) that unit nonresponse significantly decreases with income. As non-western

immigrants are often in the lower part of the income distribution this may play a role⁵. On the other hand Feskens et al. (2007) found that the effect of ethnicity on response is almost entirely mediated by the degree of urbanization. Urbanization has a negative effect on the contact probabilities, especially for sample units with a non-western background. This research was based on the Continuous Survey on Living conditions (POLs) 1998.

Table 7. Ethnicity in LISS, compared with the national statistics

Origin, age>=16	Autochthonous %	1st generation immigrant %	2nd generation immigrant %	Total %
LISS				
Autochthonous	87.7	0.0	0.0	87.7
Non-western immigrant	0.0	3.3	1.2	4.5
Western immigrant	0.0	3.0	4.8	7.8
Total	87.7	6.3	6.0	100.0
CBS 2008 (population)				
Autochthonous	81.1	0.0	0.0	81.1
Non-western immigrant	0.0	7.3	2.2	9.5
Western immigrant	0.0	4.3	5.1	9.4
Total	81.1	11.5	7.4	100.0

Table 8. Ethnicity and gender in LISS, compared with the national statistics

Origin, age>=16	LISS		CBS 2008 (population)	
	male %	female %	male %	female %
Autochthonous	40.9	46.8	39.8	41.3
Non-western immigrant	1.9	2.6	4.8	4.7
Western immigrant	3.6	4.2	4.4	5.0
Total	46.4	53.6	49.0	51.0
Autochthonous	40.9	46.8	39.8	41.3
1st generation immigrant	2.7	3.6	5.5	6.0
2nd generation immigrant	2.8	3.2	3.7	3.7
Total	46.4	53.6	49.0	51.0

Table 9. Ethnicity in LISS, compared with the national statistics

Origin, age>=16	LISS	CBS 2008 (population)
	%	%
Dutch	87.7	81.1
Turkish	0.9	2.0
Moroccan	0.6	1.7
Netherlands Antilles	0.8	0.7

⁵ In this version of the paper we have not included income in our analysis. In the LISS panel income of the household members has been asked in the general household questionnaire, which is answered by one of the household members. Furthermore, in June there has been an individual questionnaire on income and wealth for all household members age 16 and older. Unfortunately, not all respondents have answered these questions, and probably the non-response is not random over the income distribution. In a later stage we will focus on income and the non-response related to these questions.

Surinam	0.8	2.0
Indonesia	2.5	2.8
Other non-western	1.4	3.1
Other western	5.3	6.6
Total	100	100

Table 10. Ethnicity and age groups in LISS, compared with the national statistics

Age	Autochthonous %	Non-western immigrant %	Western immigrant %	Total %
LISS				
16-19	5.4	0.5	0.5	6.3
20-39	25.9	2.2	2.4	30.5
40-64	45.7	1.7	3.7	51.1
65-79	9.8	0.1	1.2	11.1
80+	0.9	0.0	0.1	1.1
Total	87.7	4.5	7.8	100.0
CBS 2008 (population)				
16-19	4.7	0.9	0.4	6.1
20-39	24.2	4.8	3.1	32.2
40-64	36.2	3.3	4.1	43.6
65-79	11.8	0.4	1.4	13.6
80+	4.2	0.0	0.4	4.6
Total	81.1	9.5	9.4	100.0

Having examined the presence of different ethnicities in LISS, we will now continue investigating the representativeness of the people who answered the REE and HEALTH questionnaires. This is important in case non-response is nonrandom. When, for example, especially the non-western immigrants have not answered the REE questionnaire, it may be the case that the complete LISS panel is more representative with regard to ethnicity than is appearing from the above analysis (or the other way around). On the other hand, when these people and other people systematically do not answer individual questionnaires, one may still question whether research results may be considered to be representative for the population.

We investigate the representativeness of the people who answered the REE and HEALTH questionnaire by comparing the background characteristics of these people with the national statistics. In this way we can find out whether people with specific characteristics have not answered these two individual questionnaires. Individual questionnaires are only presented to household members of age 16 and older, therefore, the national statistics also only contain the people of age 16 and older. The HEALTH questionnaire was collected during November 2007 and February 2008, the questionnaire about religion and ethnicity was collected during January 2008 and April

2008. 9,845 of the 12,626 persons of the LISS panel are eligible for the individual questionnaires (they have the age of 16 or older).

Table 11 shows the response of these people to the REE and HEALTH questionnaires. About 75% responded to the REE questionnaire, and about 67% to the questionnaire about health. 64% has answered both questionnaires.

Table 11. Number of respondents and non-respondents in REE and HEALTH

	not responded to HEALTH	responded to HEALTH	Total
not responded to REE	2,191	318	2,509
responded to REE	1,045	6,291	7,336
Total	3,236	6,609	9,845

Below we present the characteristics of the people who answered the individual questionnaires, and compare them with national statistics. Table 12 shows that although the young (age 16-20) are overrepresented in the LISS panel, their response to the individual questionnaires is relatively low. Therefore, in the individual questionnaires their overrepresentation is diminished. Between the age of 16-64 we see that females have higher response rates than men in the individual questionnaires (given the fact that they are present in the LISS panel). This causes women to be somewhat overrepresented in the two individual questionnaires. For the persons aged 65 and older the situation is different. There men have a higher response rate than women in the individual questionnaires, possibly at least partly because they are more familiar with computers. REE has a higher response rate than HEALTH, especially for the young (16-20) and the elderly (80+).

Table 12. Age and gender in LISS, the individual questionnaires, and the national statistics

Age	LISS		REE		HEALTH		CBS 2007 (private households)	
	Male	Female	Male	Female	Male	Female	Male	Female
16-20	3.7	3.6	3.0	3.3	2.6	2.9	3.1	3.0
20-39	15.5	16.8	13.3	17.3	13.1	17.7	16.6	16.5
40-64	24.5	25.0	23.6	27.4	24.1	27.7	22.0	21.7
65-79	5.2	4.8	5.8	5.2	5.9	5.1	6.2	7.2
80+	0.6	0.4	0.6	0.4	0.6	0.3	1.4	2.5
Total	49.4	50.6	46.4	53.6	46.2	53.8	49.2	50.8

Table 13 and Table 14 show the proportions of persons subdivided according to marital status in LISS, the individual questionnaires, and the national statistics. In the LISS panel married people are overrepresented. In Table 13 we see that married women are also responding to the individual questionnaires relatively well, so that they are there even more overrepresented there. Married men are answering the individual questionnaires to a lesser extent. As a result, their overrepresentation in the individual questionnaires is smaller. Never married men are underrepresented in the LISS panel. Given that they are in the LISS panel, they also have a relatively low response rate to the individual

questionnaires, so that this group is even more underrepresented in the individual questionnaires. Divorced people are a little bit underrepresented in the LISS panel, however, their response to the individual questionnaires is relatively high, and they are not underrepresented any more in the two individual questionnaires.

Table 13. Marital status in LISS, the individual questionnaires, and the national statistics

Marital status	LISS		HEALTH		REE		CBS 2007 (population)	
	Male	Female	Male	Female	Male	Female	Male	Female
Married	29.1	29.6	28.4	32.3	28.1	31.8	26.3	26.2
Divorced	3.1	4.3	3.3	4.7	3.4	4.6	3.3	4.4
Widow(er)	0.7	1.9	0.8	2.1	0.8	2.2	1.3	5.3
Never married	16.5	14.8	13.7	14.7	14.1	15.0	18.2	15.0
Total	49.4	50.6	46.2	53.8	46.4	53.6	49.0	51.0

Table 14. Marital status and age in LISS, the individual questionnaires, and the national statistics

LISS					
Age	Married	Divorced	Widow(er)	Never married	Total
16-20	0.0	0.0	0.0	7.3	7.3
20-39	12.7	1.2	0.0	18.3	32.2
40-64	37.5	5.5	1.0	5.4	49.4
65-79	7.9	0.7	1.2	0.3	10.0
80+	0.6	0.0	0.4	0.0	1.0
Total	58.6	7.4	2.6	31.3	100.0

REE					
Age	Married	Divorced	Widow(er)	Never married	Total
16-20	0.0	0.0	0.0	6.3	6.3
20-39	12.7	1.2	0.1	16.6	30.6
40-64	38.0	6.0	1.2	5.8	51.0
65-79	8.5	0.8	1.4	0.4	11.1
80+	0.6	0.0	0.4	0.0	1.1
Total	59.9	8.0	3.0	29.1	100.0

HEALTH					
Age	Married	Divorced	Widow(er)	Never married	Total
16-20	0.0	0.0	0.0	5.6	5.6
20-39	13.0	1.2	0.0	16.6	30.8
40-64	38.8	6.0	1.2	5.8	51.8
65-79	8.5	0.8	1.3	0.3	10.9
80+	0.5	0.0	0.4	0.0	0.9
Total	60.7	8.0	2.9	28.4	100.0

CBS 2007 (population)					
Age	Married	Divorced	Widow(er)	Never married	Total
16-20	0.0	0.0	0.0	6.0	6.0
20-39	11.5	1.2	0.0	20.1	32.7

40-64	30.8	5.3	1.2	6.1	43.3
65-79	8.8	1.1	2.8	0.7	13.4
80+	1.4	0.2	2.6	0.3	4.6
Total	52.5	7.7	6.6	33.2	100.0

Table 15 shows the cross tabulation of age and the region of residency for the LISS panel, the individual questionnaires, and the national statistics. We already saw that the age group 20-44 is overrepresented in the LISS panel in each region except for the west. Here we see that their response rate with respect to the individual questionnaires is also relatively low. As a result, they are somewhat more underrepresented in the individual questionnaires. We do not find notable differences between the response to the REE and HEALTH questionnaire in this respect.

Table 15. Region of residence in LISS, the individual questionnaires, and the national statistics

LISS					
Age	North	East	South	West	Total
20-44	5.1	10.3	10.6	20.7	46.6
45-64	4.6	9.1	10.7	17.1	41.5
65-79	1.2	2.6	2.5	4.6	10.8
80+	0.2	0.1	0.3	0.5	1.1
Total	11.0	22.1	24.0	42.9	100.0
REE					
Age	North	East	South	West	Total
20-44	4.8	9.9	10.4	19.1	44.2
45-64	4.6	9.3	11.1	17.8	42.9
65-79	1.1	2.7	2.8	5.2	11.8
80+	0.2	0.1	0.3	0.5	1.1
Total	10.7	22.1	24.6	42.6	100.0
HEALTH					
Age	North	East	South	West	Total
20-44	4.8	9.6	10.5	19.3	44.2
45-64	4.7	9.4	11.5	17.7	43.2
65-79	1.1	2.7	2.8	4.9	11.6
80+	0.2	0.2	0.2	0.4	1.0
Total	10.8	21.8	25.0	42.4	100.0
CBS 2008 (population)					
Age	North	East	South	West	Total
20-44	4.5	9.3	9.2	21.6	44.7
45-64	3.8	7.5	8.2	16.4	35.9
65-79	1.6	3.0	3.4	6.4	14.4
80+	0.6	1.0	1.0	2.4	4.9
Total	10.4	20.8	21.9	46.8	100.0

Note that in this table the youngest age group starts at age 20 instead of 16.

5. Multivariate analysis of response

As we saw in section 4, the composition of the set of respondents to the questionnaire on religion and ethnicity differs to some extent from the composition of the full Liss panel. In some cases, the underrepresentation of a group is larger, in other cases it is smaller. In this section we want to answer the question which characteristics can be seen to be associated with overrepresentation and underrepresentation in the data on religion and ethnicity (REE), keeping all other variables constant. In this case, we are able to perform a multivariate analysis including all variables of interest in the equation at the same time. Unfortunately, because the ethnic origin of the respondent is one of the topics of the questionnaire, we cannot infer whether panel members with various ethnic origins are more or less likely to respond to this questionnaire.

The explanatory variables used in the multivariate analysis include: household size, gender, age and age squared, the cross-product of gender and age and age squared, the position in the household, the extent of urbanisation, home-ownership, whether or not the respondent is responsible for the household accounts, the region of residency, the type of household, the education level completed, and the activity status of the respondent. In general, the age effect turns out to be quite different between males and females. In most other cases, the interaction terms with gender turn out to insignificant, with the exception of the interaction between gender and working in the family business. In our preferred specification we only include these significant interaction terms.

Next to the response to the religion and ethnicity questionnaire we also present a multivariate analysis of the response to the questionnaire on health. The results of both analyses turn out to be quite comparable. To some extent this could be expected because in both cases part of the non-response consists of members of panel households who have chosen not to participate in the panel at all. The households in question remain in the panel when other (adult) household members do participate. In other words, parts of the nonresponse of both questionnaires was known to coincide beforehand. To get a better understanding of the differences between the response to both questionnaires, we will disentangle non-participation of household members from non-response, and will show results on non-participation and on non-response of participants separately. In doing so, we can show which variables can be shown to be correlated with differences in participation and which variables are correlated with differences in response. In some cases, variables have similar effects in both phenomena, in other cases, variables are associated with non-participation do not have a differentiating effect on non-response and vice versa.

Table 16 shows the results of a logit analysis of the response to the questionnaires on religion and ethnicity and health. Although the (pseudo-)r² of the equations in question is not very high, quite a few explanatory variables have significant coefficients. The probability of filling in the questionnaire on religion and ethnicity decreases slightly with household size. Compared to heads of household, unmarried partners and other household members (not being partner or children of the head) have a lower probability of responding to this questionnaire. There are no significant differences between areas distinguished by the level of urbanization. Tenants have a lower tendency to respond than home-owners, and persons who are responsible for the household accounts are significantly more likely to fill out the questionnaire than others. For inhabitants of the three largest cities the response probability is lower than for inhabitants of the other regions. Persons in couples without children, and especially single persons are more likely to fill out the questionnaire than persons in couples with children. There are also significant differences between persons with various levels of education. In general, persons with higher education have a higher response probability than persons with lower education, but there is no significant difference between persons with university education and lower secondary education. Compared to persons in paid employment, self-employed and especially male persons employed in the family firm are less likely to respond but pensioners and especially students have a higher response probability.

The relationship between the probability to respond to the REE questionnaire and age for males and females is presented in Figure 1. In this figure, the probabilities based on the logit analysis have been calculated differentiated by age and gender but with all other explanatory variables fixed at the average for the population in question. For males, the response probability increases with age, more so for the young than for the elderly, and reaches its maximum at about age 70. For females, the response probability is clearly higher than for males, except for the elderly. The response probability for females increases until about age 50 and decreases afterward⁶.

To a large extent, the coefficients of the logit equation representing the probability to respond to the HEALTH questionnaire are quite comparable to the coefficients discussed so far. In contrast to the REE questionnaire, the response does not decrease significantly with household size and it is not significantly lower for tenants than for homeowners. On the other hand, the response probability is lower in moderately urban and hardly urban areas, in comparison to very urban areas. A number of coefficients for response to the health questionnaire are clearly lower than for the REE questionnaire, but still suggest a

⁶ It should be taken into account that these results have been calculated keeping all other variables constant. The actual differences between males and females in the various age groups will be different as a result of differences in the other characteristics.

significant difference. In particular this holds for the coefficients for unmarried partners, single persons, students, pensioners and males employed in the family firm.

The relationship with age, differentiated by gender is depicted in Figure 2. Again, the figure shows a close resemblance to Figure 1.

As mentioned earlier, to some extent the close correspondence of the results on the response to both questionnaires could have been expected because part of the non-response consists of persons who have refused to participate in the panel in households where other (adult) persons do participate. Persons who have refused to participate were categorized as non-respondents in both logit equations presented so far. In Table 17 we present the results of logit equations of the probability to respond to both questionnaires, given that one has agreed to participate in the panel. In this case, the fact that both equations yield similar results actually allows conclusions about the response behaviour of the participants to the panel: variables associated with non-response to the first questionnaire are also associated with non-response to the second.

Response decreases with household size in both cases, and for unmarried partners we find a significant negative coefficient for the religion and ethnicity questionnaire only. In contrast to the results of Table 16, we find a significantly positive coefficient for children. However, as can be inferred from Table 18, children have a negative coefficient in a logit equation of the probability to participate. In other words, children are less likely to participate, but, once they participate, they are more likely to respond. On the other hand, other household members (not being head, spouse or child) are less likely to participate, and if they participate, they are also less likely to respond. Another noteworthy result is that being responsible for the household accounts does not affect the response behaviour of the participants, but it does increase the likelihood of participation. Likewise, in contrast to Table 16, in Table 17 we do not find a significant coefficient for single persons. This can be seen to be caused by the fact that all single persons (in fact, all adults in households with one adult) in the panel are participating (since if not, their household would drop out of the panel). Obviously, these persons are not included in the logit analysis of participation reported in Table 18.

When they participate, members of single parent families are less likely to respond to both questionnaires than members of the reference household type, couples with children. The education coefficients of Table 17 are by and large comparable to Table 16 and the coefficients of Table 18 show the same general tendency. In other words, the education groups more likely to participate are also more likely to respond, given that

they participate. Compared to persons in paid employment, once again we find that males employed in the family firm and self-employed are less likely to respond to the REE questionnaire, whilst students and pensioners show significant positive coefficients. As regards the response to the health questionnaire, both males and females working in the family firm, and self-employed show a lower response probability than the reference group of persons in paid employment, but the coefficients of students and pensioners are insignificant. The probability to participate (Table 18) is lower for self-employed and male employees in the family firm, and significantly higher for students than for those in paid employment.

Figure 3 and Figure 4 present the age effects of males and females on the probability of participants to respond to both questionnaires. Again the figures are fairly comparable. In comparison to Figure 1 and Figure 2, we see that the difference between males and females is clearly smaller. The probability of participants to respond to the questionnaire on religion and ethnicity increases by age until about age 60 and hardly decreases afterward. The probability of participants to respond to the health questionnaire increases until about age 60 for males and until about age 50 for females. Especially for females, the probability clearly decreases in the higher age groups.

As to the probability to participate, Figure 5 shows that except for the elderly, females participate more. However, the probability to participate does not increase by age. In fact, for males it shows a decreasing tendency until about age 60, with a slight increase in the higher age groups, and for females, the participation probability is highest between age 20 and 40 and shows a clear decrease afterward. Notably, it should be kept in mind that the estimates of Table 18 and Figure 5 are based on households with more than one adult only. Therefore, they are not completely comparable with the other tables and figures.

6. Conclusion

In a number of respects, the composition of the LISS panel differs significantly from the Dutch population. Although exact statistics about the target population are not available we can conclude that there are differences by a.o. household size, age, education, marital status, gender, and level of urbanization. Moreover, the composition of the respondents to separate questionnaires differs from that of the full panel. In part, this is caused by individual (adult) household members who do not want to participate in the panel, but also by participating household members who do not fill out individual questionnaires. When we concentrate on the response of participants, a multivariate

analysis shows that there is considerable overlap between the variables associated with differences in the response probability of two separate questionnaires.

As far as differences with respect to ethnic composition are concerned, we can only draw conclusions for the respondents to the questionnaire on religion and ethnicity. Here, we find that the ethnic composition differs from the total Dutch population. To some extent, this may be caused by the fact that respondents have to have sufficient command over the Dutch language. However, this is unlikely to be the full explanation for the difference.

The differences found suggest that unweighted estimates based on the data from the separate questionnaires may not be completely unbiased. Whether and how weighting could decrease this bias is beyond the scope of this paper. Although the composition of the set of all participants to the panel also deviates from the target population, it would still be advisable that all attempts should be made to increase the response to individual questionnaires. Moreover, as far as possible, deterioration of the representativeness as a result of panel attrition should be avoided.

Notably, to some extent, the results of this paper should be considered preliminary. For instance, the logit estimates are not based on explicit theories of participation and response but include almost all available explanatory variables. We do not take into account possible endogeneity and do not exploit the panel nature of the data but estimate all equations separately. One of the directions for further research will be to estimate more sophisticated and theoretically well-founded, possibly simultaneous models of participation and response in which we will also pay explicit attention to the correlation between the response to different questionnaires. Nevertheless, even the fairly simplistic approach of this paper shows convincingly that participation and response do differ significantly between persons with different characteristics.

7. References

Bethlehem, J. (2006), 'Representativity of web surveys – an illusion?', in Stoop en Wittenberg (2006), pp. 19-44.

Biemer, P.P. and S. L. Christ (2008), Weighting survey data, in: E.D. de Leeuw, J.J. Hox and D.A. Dillman (eds), *International Handbook of Survey Methodology*, Lawrence Erlbaum Ass., New York, pp. 317-341.

Feskens R. , J. Hox, G. Lensvelt-Mulders and H. Schmeets (2007), Nonresponse among ethnic minorities: a multivariate analysis. *Journal of Official Statistics*, 23, pp. 387-408.

Groves, R.M., F.J. Fowler jr, M.P. Couper, J.M Lepkowski, E. Singer and R. Tourangeau (2004), *Survey methodology*, Wiley, Hoboken, N.J.

Kalwij A. (2007). An empirical analysis of the relationship between income and unit nonresponse in the survey of health, ageing and retirement in Europe.

SCP (2007). *Jaarrapport Integratie 2007*. Sociaal en Cultureel Planbureau, Den Haag, November 2007.

Stoop, I. (2006), 'Access panels and online surveys: mystifications and misunderstandings', in: Stoop en Wittenberg (2006), pp. 5-17

Stoop, I. en M. Wittenberg (eds., 2006), *Access panels and online research, panacea or pitfall?*, DANS Symposium Publications 4, Aksant, Amsterdam

Table 16. Logit analysis of response to two questionnaires

Response	Religion etc		Health	
	coeff.	s.e.	coeff.	s.e.
Household size	-0.0657	0.0328	-0.0599	0.0314
Gender: male	0.0369	0.3199	-0.0435	0.2989
Age/10	0.7852	0.1523	0.8576	0.1388
(Age/10)**2	-0.0817	0.0162	-0.0916	0.0147
Male*Age/10	-0.3722	0.1572	-0.3104	0.1448
Male*(Age/10)**2	0.0519	0.0173	0.0475	0.0157
Position: Married partner	-0.0026	0.0883	0.0613	0.0808
Position: Unmarried partner	-0.3589	0.1078	-0.2397	0.1018
Position: Child	-0.1878	0.1324	-0.0507	0.1249
Position: Other household member	-0.9800	0.2418	-1.0496	0.2465
Extremely urban (≥ 2500 addr /km ²)	-0.0469	0.0938	-0.1439	0.0865
Moderately urban (1000-1500 addr /km ²)	-0.0525	0.0710	-0.1461	0.0654
Hardly urban (500-1000 addr /km ²)	-0.1029	0.0722	-0.1715	0.0666
Not urban (< 500 addr /km ²)	-0.0350	0.0832	-0.0917	0.0768
Tenant	-0.1311	0.0625	-0.0149	0.0578
Does household accounts	0.3207	0.0563	0.2685	0.0518
Amsterdam, Rotterdam, The Hague	-0.2147	0.0861	-0.1925	0.0797
North	-0.0259	0.0889	0.0267	0.0824
East	0.0256	0.0705	-0.0277	0.0647
South	0.0939	0.0696	0.1247	0.0641
Single person	0.7390	0.1461	0.4103	0.1312
Couple without children	0.2647	0.0895	0.2004	0.0840
Single parent	0.0007	0.1211	-0.0706	0.1143
Other household type	0.1698	0.2934	0.3678	0.2890
Primary education	-0.2763	0.1101	-0.3121	0.1032
Secondary education: havo/vwo	0.3080	0.0943	0.2250	0.0855
Secondary education: mbo	0.2983	0.0709	0.2806	0.0655
Tertiary education: hbo	0.3390	0.0758	0.3641	0.0695
Tertiary education: university	0.0525	0.1005	0.1323	0.0940
Other education	-0.1114	0.1297	-0.1203	0.1205
No education completed	-0.0414	0.1327	-0.1200	0.1256
Employed in family firm	0.2110	0.2758	-0.0894	0.2311
Self-employed	-0.4792	0.0909	-0.4048	0.0873
Unemployed	0.0011	0.1898	-0.1611	0.1731
Student	0.7599	0.1209	0.4880	0.1135
Homemaker	0.1385	0.1007	0.0621	0.0909
Pensioner	0.4170	0.1302	0.2834	0.1137
Disabled	-0.0170	0.1464	-0.0074	0.1338
Other	-0.2227	0.1532	-0.1329	0.1432
Male*employed in family firm	-1.4357	0.3568	-0.8885	0.3229
Intercept	-0.5062	0.3728	-0.9589	0.3445

Table 17. Logit analysis of response of household members participating in the Liss panel

Response, given participation	Religion etc		Health	
	coeff.	s.e.	coeff.	s.e.
Household size	-0.1129	0.0424	-0.0858	0.0370
Gender: male	0.0000		0.0000	
Age/10	0.9215	0.1806	0.9760	0.1419
(Age/10)**2	-0.0782	0.0200	-0.0935	0.0154
Male*Age/10	-0.1558	0.0577	-0.1701	0.0470
Male*(Age/10)**2	0.0250	0.0105	0.0287	0.0081
Position: Married partner	0.0067	0.1125	0.0744	0.0932
Position: Unmarried partner	-0.3131	0.1368	-0.1489	0.1199
Position: Child	0.3467	0.1741	0.3516	0.1498
Position: Other household member	-0.6625	0.2993	-0.7981	0.2799
Extremely urban (>=2500 addr /km ²)	0.0543	0.1221	-0.0977	0.1019
Moderately urban (1000-1500 addr /km ²)	0.0204	0.0938	-0.1237	0.0775
Hardly urban (500-1000 addr /km ²)	-0.0371	0.0953	-0.1317	0.0792
Not urban (< 500 addr /km ²)	0.0868	0.1114	-0.0128	0.0921
Tenant	-0.2827	0.0794	-0.0657	0.0672
Does household accounts	0.0778	0.0762	0.0786	0.0627
Amsterdam, Rotterdam, The Hague	-0.2465	0.1145	-0.1872	0.0951
North	-0.1321	0.1174	-0.0198	0.0983
East	-0.0738	0.0946	-0.1109	0.0771
South	-0.0480	0.0931	0.0505	0.0768
Single person	-0.1025	0.1753	-0.1760	0.1483
Couple without children	0.1054	0.1190	0.0409	0.1003
Single parent	-0.2882	0.1490	-0.2884	0.1300
Other household type	-0.3174	0.3516	0.0936	0.3290
Primary education	-0.3350	0.1458	-0.3315	0.1214
Secondary education: havo/vwo	0.2639	0.1247	0.1576	0.1004
Secondary education: mbo	0.2434	0.0944	0.2221	0.0784
Tertiary education: hbo	0.2825	0.1020	0.3201	0.0838
Tertiary education: university	0.0648	0.1346	0.1709	0.1144
Other education	-0.0240	0.1816	-0.0584	0.1479
No education completed	-0.0971	0.1661	-0.1846	0.1422
Employed in family firm	0.0460	0.3316	-0.4321	0.1920
Self-employed	-0.3647	0.1253	-0.2750	0.1093
Unemployed	-0.0857	0.2356	-0.2348	0.1972
Student	0.5113	0.1527	0.2074	0.1316
Homemaker	0.2165	0.1347	0.0580	0.1078
Pensioner	0.5861	0.2015	0.2639	0.1406
Disabled	-0.0639	0.1852	-0.0567	0.1550
Other	-0.3036	0.1921	-0.1474	0.1664
Male*employed in family firm	-1.2049	0.4482		
Intercept	-0.1889	0.4469	-0.7533	0.3665

Table 18. Logit analysis of participation (members of households with two or more adults)

Participation	coeff.	s.e.
Household size	0.0091	0.0408
Gender: male	0.2393	0.4243
Age/10	0.3882	0.2061
(Age/10)**2	-0.0665	0.0214
Male*Age/10	-0.6854	0.2073
Male*(Age/10)**2	0.0916	0.0224
Position: Married partner	-0.0284	0.1206
Position: Unmarried partner	-0.4311	0.1413
Position: Child	-0.9513	0.1697
Position: Other household member	-1.2272	0.2922
Extremely urban (≥ 2500 addr /km ²)	-0.1474	0.1232
Moderately urban (1000-1500 addr /km ²)	-0.1325	0.0925
Hardly urban (500-1000 addr /km ²)	-0.1818	0.0941
Not urban (< 500 addr /km ²)	-0.1575	0.1069
Tenant	0.1143	0.0853
Does household accounts	0.5227	0.0731
Amsterdam, Rotterdam, The Hague	-0.1462	0.1109
North	0.1185	0.1143
East	0.1767	0.0903
South	0.2531	0.0897
Couple without children	0.4970	0.1130
Single parent	0.2904	0.1789
Other household type	0.8179	0.3953
Primary education	-0.1765	0.1391
Secondary education: havo/vwo	0.3265	0.1227
Secondary education: mbo	0.3388	0.0910
Tertiary education: hbo	0.3440	0.0972
Tertiary education: university	0.0562	0.1283
Other education	-0.1722	0.1594
No education completed	-0.0040	0.1756
Employed in family firm	0.4485	0.4328
Self-employed	-0.5099	0.1095
Unemployed	-0.0365	0.2507
Student	0.9513	0.1553
Homemaker	0.0084	0.1335
Pensioner	0.1956	0.1572
Disabled	0.0786	0.2056
Other	-0.1570	0.2046
Male*employed in family firm	-1.3834	0.4894
Intercept	1.4736	0.5081

Figure 1.

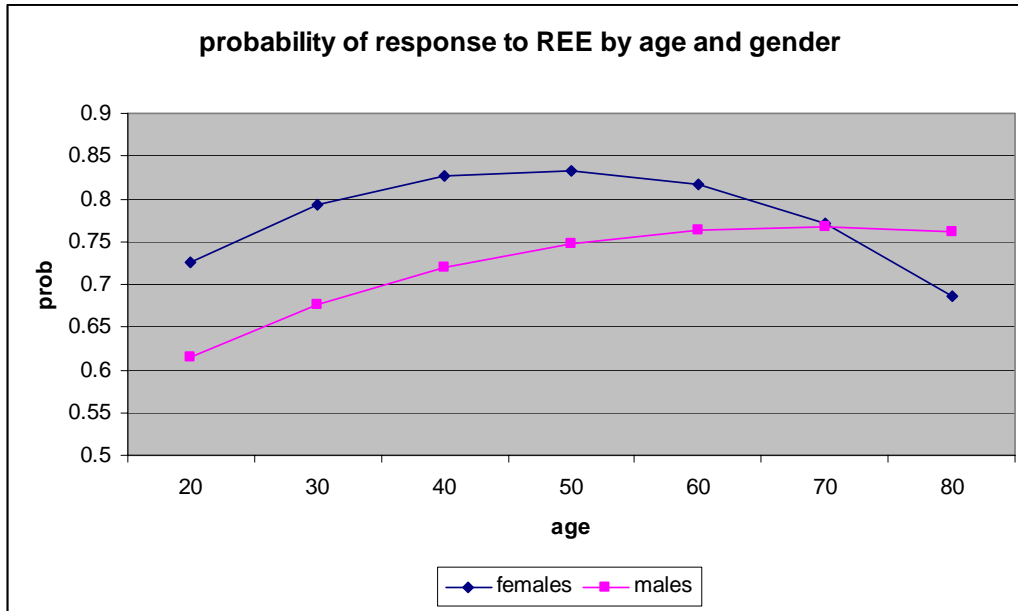


Figure 2.

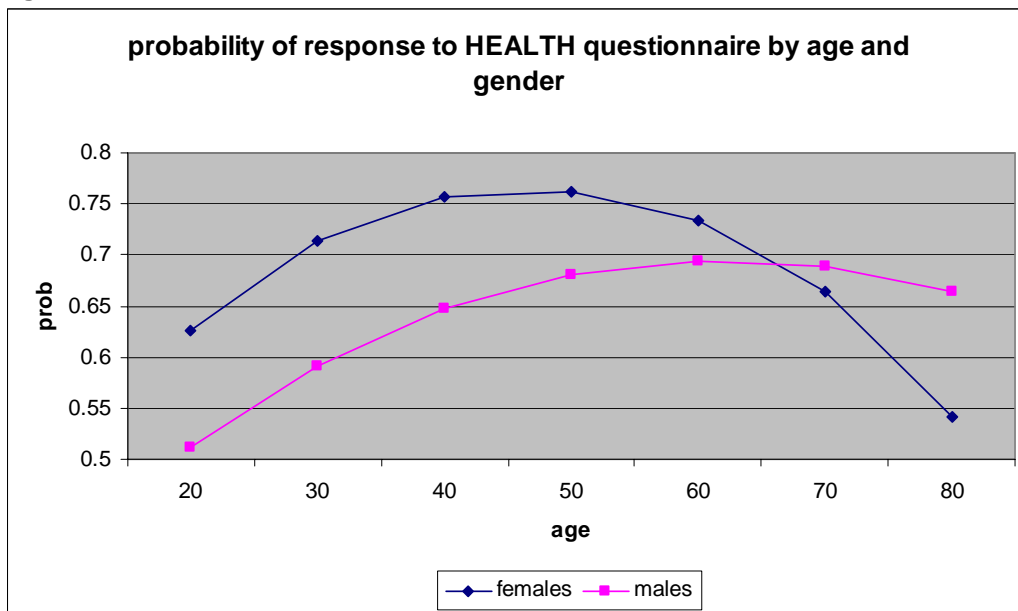


Figure 3.

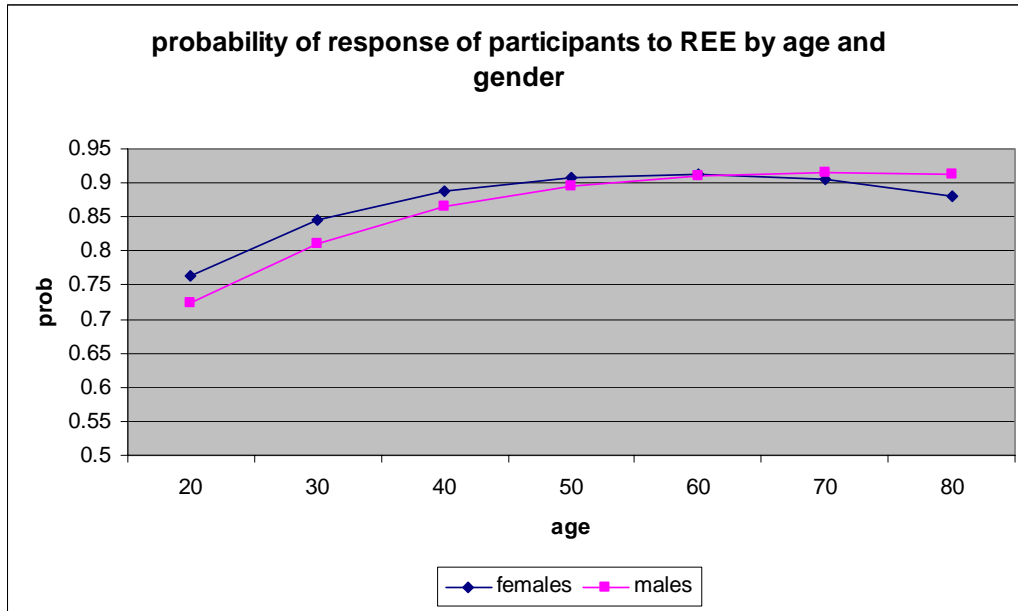


Figure 4.

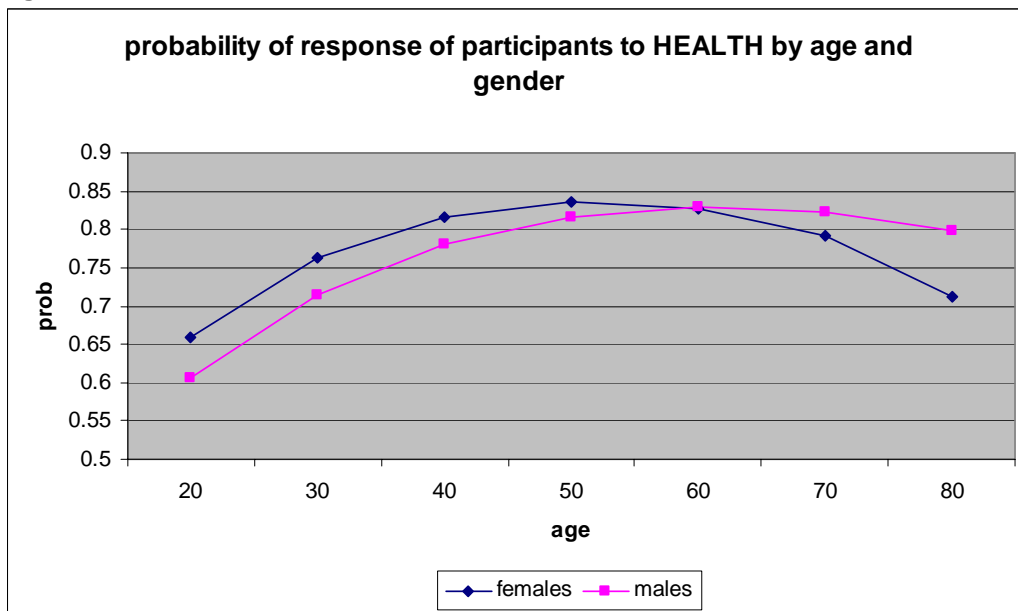


Figure 5.

