International Journal of Internet Science
2013, 8 (1), 17–29 ISSN 1662-5544

International Journal of Internet Science

IJIS.NET

Does It Pay Off to Include Non-Internet Households in an Internet Panel?

J. Leenheer, A. C. Scherpenzeel

CentERdata, Tilburg University, The Netherlands

Abstract: This paper investigates whether it is possible to improve the representativeness of an Internet panel by including non-Internet households. We study the LISS panel, managed by CentERdata, an Internet panel based on a probability sample that comprises approximately 5000 households. The LISS panel provides non-Internet households, households with no Internet access at the time of the sampling, with cost-free equipment and an Internet connection. Early 2010 the LISS panel contained 545 non-Internet households, this equals approximately 10% of the entire panel.
The analyses show that particularly older households, non-western immigrants, and one-person households are less likely to have Internet access. The LISS panel includes a representative sample of non-Internet households except for households with high average age ("the oldest old"). Non-Internet households who participate in the panel show higher response rates on the individual questionnaires and lower attrition rates. While significant differences between the panel and the Dutch population remain, the complete LISS panel, with both Internet and non-Internet households, appears to be closer to the Dutch population than the panel consisting only of Internet households for all socio-demographic variables we tested. Furthermore, about half of the non-Internet households start to use the Internet after they have become panel members. They use less of the options offered by Internet, and mainly use the simpler applications, such as e-mail and information search, compared to persons living in Internet households. In this sense, they remain different from the original Internet households and continue to contribute to the quality of the panel data.

Keywords: Online-panels, panel representativeness, Internet adoption

Introduction

Internet surveys are nowadays widely used for socio-economic research, market research, and opinion polling. They often provide the quickest and most cost-effective way to gather data. As a result of the declining number of households with a landline telephone, and the corresponding decrease in landline coverage, telephone surveys are losing popularity. A further advantage of Internet surveys is that they eliminate the possibility of response bias caused by the presence of an interviewer ('interviewer effects'). In addition, Internet surveys can be interactive to some degree, unlike, for instance, written surveys. One may for example include videos or visual applications, and make use of complex routing or preloaded files. On the other hand, one of the greatest challenges that Internet surveys face is covering the entire population. Although the penetration of the Internet has increased worldwide throughout the years, it varies widely among countries. For example, in 2011 Internet penetration measured 78% in North America, 49% in Italy, 70% in France, 80% in Germany, 88% in the Netherlands, and 92% in Sweden

Address correspondence to Jorna Leenheer. CentERdata, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: j.leenheer@uvt.nl. Tel: +31 13 488 3597. Fax: +31 13 466 2764.

(www.Internetworldstats.com). In Europe as a whole, the Internet penetration rate is 58%, and worldwide it stands at 30%. At household level, the Internet penetration rate in the Netherlands stands at 92% (www.statline.nl).

In countries with low Internet penetration, the coverage of Internet panels is obviously an issue. However, even in countries with high Internet penetration, like the Netherlands, the representativeness of Internet surveys remains a point of concern (Couper, 2000). As Internet penetration in a country expands, households without Internet access form an increasingly atypical subgroup that differs increasingly from households with Internet access. This phenomenon is known as the "digital divide" (Mahajan, Muller, & Srivastava, 1990) or the "Couper paradox" (Couper, 2000).

There are different ways of dealing with the incomplete coverage of Internet panels and the resulting problems of representativeness. Two commonly used solutions are survey weighting and mixed-mode designs. However, several studies have shown that weighting data based on socio-demographic factors does not in all cases solve the issue of representativeness with regard to non-Internet households (Couper, Kapteyn, Schonlau, & Winter, 2007; Zhang, Callegaro, & Thomas, 2008; Rookey, Hanway, & Dillman, 2008; Loosveldt & Sonck, 2008). Weighting does not correct for differences in the properties and characteristics of households with and households without Internet access *within* socio-demographic groups (Brown & Venkatesh, 2005). These differences can include lifestyle characteristics such as Internet use, which are relevant variables in market and socio-economic research and often correlated with the research matter. Using mixed-mode designs in which households without Internet access are approached through another medium, such as telephone or regular mail, may provide an alternative. Dillman, Reips, and Matzat (2010) list five ways to effectively survey the general public, based on developments in mixed-mode research in the last decade. Their list ranges from the advice to include a token cash incentive with the survey request to combining a Web survey with a mail non-response follow up, all of which have shown to lead to higher response rates and better sample representation. As one example, Rookey, Hanway, and Dillman (2008) showed that allowing households without Internet access, in a probability-based household panel, to participate in a survey by regular mail indeed improved the quality of the survey results. The downside to this approach, however, is that various features of the Internet, such as sound files and videos, cannot be used with households without Internet access. In addition, mixed-mode surveys are often influenced by so-called "mode effects" that result from the use of different interview methods (Dillman et al., 2009).

Our purpose in this article is to evaluate an alternative approach to improving the coverage of Internet panels. We examined the LISS panel ('Longitudinal Internet Studies for the Social Sciences'), a Dutch household panel based on a random sample from the Dutch population. The participants in this panel filled out various questionnaires on a monthly basis. Non-Internet households that were willing to participate in the panel were provided with access to the Internet and, if necessary, also with a computer (see www.lissdata.nl). We define non-Internet households as households that do not possess an Internet connection at the time that they are invited to join the panel. Other institutes that have built this type of Internet panel are Knowledge Networks and RAND in the United States, and more recent initiatives are the probability based Internet panel ELLIPS (Étude Longitudinal par Internet Pour les Sciences Sociales, operated by the Centre for socio-political data at Sciences Po) in France and the GIP (German Internet Panel, operated by the University of Mannheim) in Germany. Whereas these institutes all provide Internet access to the non-Internet households, the representativeness of the panels can still be decreased by variation in response probabilities between Internet households and non-Internet households. As a result of non-response selectivity in the recruitment of the non-Internet households, the gain in terms of improvement of the panel composition by the inclusion of these households might be limited. Hoogendoorn and Daalmans found considerable differences in the 'total initial response rate' for pc-ownership: 16% for pc-owners versus 5% for non-pc-owners. They conclude that in the panel they studied, the strategy of providing non-pc-owners with the infrastructure only works partially: it is not as bad as ignoring this group–which would be equivalent to a total response rate of 0%–but the total rate of non-pc-owners is far from the 16% of pc-owners. Similarly, we found a large difference in recruitment rate between Internet households and non-Internet households in the LISS panel: 84% for Internet households and 35% for non-Internet households (see also Table 2).[1] The question rises whether the representativeness of the panel is indeed effectively increased by this approach, when selectivity bias as a result of differential non-response is likely to exist.

The research question that we address in this article is whether providing non-Internet households with a means of Internet access, which requires a substantial investment, significantly improves the quality of an Internet panel. Furthermore, we evaluate whether this investment is effective for the long-term quality of the panel. In order to perform these tests, several matters must be examined. The first is if the recruited non-Internet households in the

[1]Our registration rate is not directly comparable in size to the total initial response rate reported by Hoogendoorn and Daalmans (2009), due to differences in response rate definition and design of the recruitment steps. Despite these differences, the importance of the difference between the pc-owners or Internet household on the one hand, and the non-pc-owners or non-Internet households on the other hand is clear in both studies.

panel are representative of the subgroup of non-Internet households in the overall population, given the low response probability in this subgroup. To answer this question, we first determine to what extent non-Internet households differ from Internet households in the total sample of the Dutch population (see step 1 in the results section). Next, we compare the non-Internet households recruited into the panel to the overall contacted sample of non-Internet households (see step 2 in the results section). Many studies have shown that non-Internet households differ significantly from Internet households in demographic characteristics and attitudinal variables. Couper (2000) described that in the Current Population Survey (CPS), conducted by the U.S. Bureau of the Census, Black and Hispanic households, households with low income levels, and low education were found to be less likely to have home Internet access. In addition, significant differences in access were found by age, rural/urban status, and region of the country. In 2009, DiSogra, Mario Callegaro, and Hendarwan showed that households without home Internet access in the U.S. still tended to be lower income, lower education and more likely Hispanic. Hence, it seems logical that including such households, by providing them Internet access, will improve the composition of an online panel on these demographic variables. However, as far as we can see, none of the studies tested whether this approach resulted in a significant improvement of the representativeness of the online panel, nor did they evaluate whether the non-Internet households that were successfully recruited in the panel were a good representation of the total sample of contacted non-Internet households. These questions seem especially relevant in relationship to the low response rates of non-Internet households compared to Internet households, as given above.

The second issue concerns the non-Internet households' loyalty to the panel, which relates to the long-term benefits of the investment to provide them with Internet access. How well do they participate in the monthly questionnaires and remain panel member for an extended period of time (see step 3 in the results section)? A final relevant issue is whether the non-Internet households continue to be representative of the subgroup of the overall population of non-Internet households, after they have been provided with Internet access for the purpose of participating in the panel. It is possible that this affects these non-Internet households, causing them to grow more similar to Internet households (see step 1 in the results section).

We answer these questions through an empirical analysis of the LISS panel, using available information from the population registers concerning the sampling frame of the LISS panel, data obtained from the questionnaires completed by the panel, and data from the panel management system. The remainder of this article is structured as follows: in the next section, the LISS panel's sampling frame and the procedure by which non-Internet households are provided with a means of Internet access are explained. The results of the empirical analysis are described in the following section. The final section presents the conclusions and discusses the implications of the findings for online panels which aim to be representations of the general population.

Sample and recruitment

LISS panel

The LISS panel is an Internet panel consisting of Dutch households established in 2007, which serves primarily scientific purposes. The LISS panel is administered by CentERdata (Tilburg University, The Netherlands).

In January 2010 the panel consisted of 8'686 respondents belonging to 5'142 households. Per month this can entail several short or one or two longer questionnaires. It usually takes between 30 and 40 minutes a month to complete all questionnaires. Some of the questionnaires are so-called 'core questionnaires' that are repeated annually.

For each household there is one member that completes the data pertaining to that household, such as household composition and the ages and daily activities of the household members. This household member is regularly requested to update this information. Further, every household member aged 16 or older indicates whether he or she wishes to participate in the monthly questionnaires or not. A household is included in the LISS panel if at least one of the members aged 16 or older agrees to participate.

To leave the LISS panel, panel members have to contact the LISS panel helpdesk, by email, telephone, or by leaving a message on their panel member Web page. The helpdesk manager can then end their participation. Before doing so, the helpdesk tries to call the panel member by telephone to conduct a short exit-interview. In addition, the helpdesk employees call all panel members that have been inactive for two months or longer,[2] and try to convince those so-called "sleeping" panel members to start participating again (Scherpenzeel & Zandvliet, 2011). Panel members who have been contacted three times or more, spread over several months, and do not

[2]Panel members without a known telephone number receive an email and/or a reminder postcard.

recommence to participate are eliminated from the panel. On average, the attrition rate of the LISS panel is about 10% per year for households and about 12% per year for persons (Scherpenzeel, forthcoming).

Sampling And Panel Recruitment. The research population of the LISS panel comprises the Dutch speaking population that has the Netherlands as permanent place of residence. The unit used to draw the random sample consists of independent private households. Institutions and other forms of collective households are excluded. Households without any adult individual with a command of the Dutch language are also excluded from the research population. Dillman et al. (2010) state that providing Internet access to non-Internet households does not always eliminate coverage problems in Internet panels since they are often based on a random digit dialing (RDD) sample. As a result of decreasing landlines, RDD samples may have large coverage problems as well (Fuchs & Busse, 2009). In contrast, the LISS panel was based on random household sample from the population register and households without a known telephone number were recruited in a face-to-face contact interview. In collaboration with Statistics Netherlands, a single random sample was drawn of 10'150 addresses from the population register. Between May and December 2007, these households were invited to participate in the LISS panel. The households were approached in a traditional manner: first an announcement letter was sent, along with a brochure explaining the nature of the panel research. A 10-Euro bill was also enclosed. Preliminary research had shown that an advance compensation of ten Euros was the most effective way of boosting people's willingness to participate in the panel (Scherpenzeel & Toepoel, 2012). After the letter, respondents were approached by an interviewer in different ways, using a mixed-mode method. Households with a known telephone number were contacted by phone (CATI). The remaining households were visited by an interviewer and were thus contacted face-to-face (CAPI). For the recruitment by phone, a maximum of 15 attempts was made to establish contact, spread across several weeks with regular intervals. If a household could not be contacted after 15 attempts, this address was subsequently approached by means of face-to-face contact. For the face-to-face recruitment, the interviewer made a series of eight attempts to establish contact. This was followed a few weeks later with a second series of seven attempts. The choice for these intensive contact efforts is supported by Feskens et al. (2007) who concluded that to maximize responses of difficult to reach groups, such as ethnic minorities, enhancing the contact rate is of key importance.

Once contact had been established, the interviewer requested a 10-minute interview with the respondents. At the end of this interview, the respondent was invited to join the panel. This invitation only followed if the interview was deemed successful. If a respondent refused to conduct the initial interview, the interviewer proposed to ask just three key questions in order to obtain some basic data about the respondent. One of those questions was whether the respondent's household has access to the Internet.

Internet Facilities for Households Without Internet access. The random sample drawn from the population register includes households that lack access to the Internet. These participants were provided with a broadband connection and if necessary also with a computer. The device offered is a so-called 'SimPC', a small and easy-to-use machine. The SimPC can be operated using large 'buttons' on the screen for the most commonly used functions, and the screen is designed in a way to make it readable for elderly persons. The broadband connection makes it possible to use visual images and video material. The computer and broadband connection were delivered to the households and fully installed. If necessary the respondents were offered help at home to learn how to operate the SimPC and how to complete the questionnaires on the screen. Support by a helpdesk remained available after installation in case of any operating problems.

Results

Step 1: To what extent do non-Internet households differ from Internet households?

To establish how households without Internet access differ from households with Internet, we use register data about all households that answered the question about Internet access during the recruitment process. [3] The total percentage of respondents that indicated to have home Internet access was 77% of the total number of respondents that answered the question, either in the recruitment interview or in the central question procedure. Table 1 (left-hand column) shows the estimates of a logit-model that explains Internet access using various socio-demographic characteristics of households. The model shows that it is less likely for a one-person household ($b = -1.073$; $z = -14.00$; $p < .001$) to have Internet access at home than for multiple-person households. Further, it is less likely for households with at least one household member of non-Western origin to have Internet access than for households that are entirely of Western origin ($b = -1.245$; $z = -8.78$; $p < .001$). In addition, Internet penetration is highest for households living in municipalities with an average urbanization level (more than 1'000–1'500 addresses per km^2) ($b = 0.288$; $z = 2.41$; $p = .02$). The variable urbanization level consists of five levels (very high,

[3]The analyses that use population register data were performed by Statistics Netherlands, as the authors do not have direct access to these data.

high, moderate, low, and without urbanization), very high urbanization (more than 2'500 addresses per km^2) is included as the reference category. For age we find an inverted U-shaped relationship with a positive linear coefficient ($b = 0.064$; $z = 4.01$; $p < .001$) and a negative quadratic coefficient ($b = -0.001$; $z = -9.14$; $p < .001$); the tipping point is at 23.8 years. This means that the likelihood of having Internet access is highest for households with an average age of adult household members of 23.8 years,[4] and that this likelihood declines at an increasing rate as the average age of the household becomes higher.

In sum, we find that households consisting of elderly individuals, non-Western immigrants, one-person households and households in regions with a high or conversely a low level of urbanization more often lack Internet access and thus will participate less frequently in Internet panels.

Table 1
Likelihood of Internet access For All Households and Likelihood of Panel Participation by Households Without Internet access (Logit-Model)

	Internet access: All households		**Panel participation: Non-Internet households**	
	Coefficient	*z*	Coefficient	*z*
One-person household	−1.07	−14.0**	−0.01	−0.1
Average age adults in household	0.06	4.0**	0.05	1.9
(Average age adults in household)^2	−0.00	−9.1**	−0.00	−3.3*
Western immigrant in household	−0.09	−0.5	−0.38	−1.1
Non-Western immigrant in household	−1.25	−8.8**	−0.11	−0.4
Region with very high urbanization[a]				
Region with high urbanization	0.05	0.5	0.18	0.9
Region with moderate urbanization	0.29	2.4*	0.09	0.4
Region with low urbanization	0.06	−0.5	−0.35	−0.2
Region with very low urbanization	0.06	−0.5	0.14	0.7
Constant	1.40	3.4**	−1.09	−1.5
N	6'919		1'517	

[a]Base line.
*$p < .05$, **$p < .01$.

Step 2: Are the recruited non-Internet households representative?

Making an Internet connection available to households without Internet may solve the representativeness problem of Internet panels, yet only under the condition that a representative sample of non-Internet households can be composed. To examine whether this succeeds in practice, we estimate a logit-model for panel participation of non-Internet households for the sampling frame of the LISS panel (Table 1, right-hand column). Non-significant coefficients indicate that the non-Internet panel members are representative for the group of non-Internet households on the explanatory variables. It appears that the likelihood of participation for non-Internet household does not depend on the size, origin or urbanization level of a household (coefficients are non-significant). A significant quadratic effect emerges for age only, with a marginally significant main effect. The effect flips at 31.3 years. This means that the likelihood of participation for non-Internet households is highest for households with an average age of 31.3 years, and that this likelihood declines–at an increasing rate–to the extent that the average age of the household is higher.

The approach by the LISS panel, to provide households with Internet access if necessary, has resulted in a substantial group of non-Internet households in the panel. In January 2010 the panel contained 545 households without Internet, 732 persons living in these households are active panel members. This means that 9.5% of the entire panel consists of non-Internet households. Table 2 shows that the recruitment rate for households without Internet in the sample frame is much lower (35.1%) than for households with Internet (84.2%).

In sum, the strategy to provide households without Internet with Internet facilities results in a substantial group of participants. These participants constitute a representative random sample of the non-Internet households in terms of ethnical background (autochthonous versus Western of Non-Western immigrants), region, and household composition (single-person versus multiple-person households). The sample is not completely representative in terms of age, higher age groups without Internet are underrepresented in the panel.

[4]Since only 2.2% of the households have an average age of its adult members under 23.8 years, the curve has a decreasing shape for most of the households. The tipping point is calculated as the maximum of a parabola, that is: 0.0641163/(2*0.0013478) = 23.8.

Step 3: To what extent are non-Internet households loyal panel participants?

The non-Internet households show a high degree of loyalty after they join the panel. That is, panel members belonging to non-Internet households completed 69.8% of the questionnaires that were presented to them, which is a significantly higher percentage than for panel members from households with Internet (59.8%) (Table 2). Attrition, between April 2008 (following the completion of the first major panel recruitment campaign) and January 2010, is 9.9% for households without Internet and 13.8% for households with Internet. Attrition means that all members of a household have decided to discontinue their participation in the panel.

Table 2
Participation and Attrition of the Households

	Internet households	Non-Internet households	Sig.	Total
Recruitment rate[a]	84.2%	35.1%	**	75.6%
Average monthly response rate (indiv.)[b]	59.8%	69.8%	**	60.5%
Attrition rate[c]	13.8%	9.9%	*	13.5%

[a] Households registered in the panel as a percentage of the total number of households that completed the recruitment interview or that answered the key question about Internet use. [b]Percentage of months in which panel member i participated (when selected), averaged over all panel members. [c]Attrition of households between 04–2008 and 01–2010.
*p < .05, **p < .01 (Chi-square/ t-tests); AAPOR figures (see Callegaro & Disogra, 2009).

Table 3
Likelihood That a Household Leaves the Panel (Logit-Model[a])

	All households		Internet households		Non-Internet households	
	Coeff.	*z*	Coeff.	*z*	Coeff.	*z*
Internet household	0.775	15.9**				
One-person household	0.005	0.002	−0.100	0.7	0.980	4.6*
Average age of adults in the household	0.017	27.9**	0.018**	29.6**	−0.001	0.01
Western immigrant in the household	−0.195	1.2	−0.198	1.1	0.096	0.01
Non-Western immigrant in the household	−0.390	3.8*	−0.383	3.4	−0.618	0.6
Homeowner	−0.007	0.004	0.018	0.03	−0.210	0.2
Net household income	> −0.001	0.06	> −0.001	0.1	> −0.001	1.4
Urbanization region:						
Very high urbanization level[b]						
high urbanization level	0.094	0.4	0.092	0.4	−0.011	< 0.001
moderate urbanization level	0.037	0.1	0.037	0.1	0.024	0.002
low urbanization level	−0.043	0.1	−0.051	0.1	−0.180	0.1
without urbanization	−0.008	0.002	0.034	0.04	−1.066	1.5
Highest educational degree in household:						
Low level (primary/VMBO)	−0.136	1.5	0.105	0.8	0.516	0.9
Medium level[b] (HAVO/VWO/MBO)						
High education (HBO/University)	0.009	0.006	−0.004	< 0.001	−0.272	0.1
Constant	−3.527	148.1**	−2.901	172.2**	−1.584	2.8
N	4’968		4’535		413	

[a] Between 04–2008 en 01–2010. [b] Base level.
*$p < .05$, **$p < .01$.

In Table 3 we show whether specific subgroups show higher attrition, between April 2008 and January 2010, using a logit-model. In line with descriptive statistics of Table 2, we find that Internet households are more likely to leave the panel than non-Internet panels ($b = 0.775$; $z = 15.9$; $p < .001$). Furthermore non-western households are less likely to leave the panel ($b = −0.390$; $z = 27.9$; $p < .001$). Overall households are more likely to leave the panel when the average age of the adults in the households is higher ($b = 0.017$; $z = 27.9$; $p < .001$). When estimating separate models for Internet and non-Internet households we find that the likelihood of leaving the panel increases

when the adult household members are older ($b = 0.019$; $z = 33.1$; $p < .001$), but for non-Internet households we did not find a relationship between age and attrition. We also find that one-person households without Internet are more likely to leave the panel than non-Internet households that consist of several people ($b = 0.980$; $z = 4.6$; $p = .04$). We do not find any relation between level of urbanization, home ownership, income or education and attrition, neither for households with Internet or for households without. To summarize, non-Internet households are less likely to leave the panel. This is even more true for one-person non-Internet households.

Representativeness of the LISS panel

An important goal of including non-Internet households in the panel is to make the resulting online panel representative for the Dutch population. Several studies exists of the representativeness in general of online panels based on probability panels. Hoogendoorn and Daalmans (2009) find non-response selectivity with respect to age and income, in a panel which started with a sample of households that had a registered telephone number. DiSogra, Callegaro, and Hendarwan (2009) found an underrepresentation of lower educated households and cell-phone only households. The LISS panel had some selectivity biases, after the first recruitment, similar to those found by Hoogendoorn and Daalmans (2009). In two studies of the representativeness of the LISS panel, Knoef and de Vos (2009), and Van der Laan (2009) found selectivity with respect to household size, age, education, marital status, gender, level of urbanization, and (initially) having a computer and Internet. These biases were partly corrected by a stratified refreshment sample, which was recruited in 2009 (Scherpenzeel, 2010). Scherpenzeel and Bethlehem (2011) compared the composition of the LISS panel to a traditional face-to-face survey, to a large volunteer based

Table 4
Representativeness of the LISS Panel With and Without Non-Internet Households, Compared to the Dutch Population

Variable	Panel with only Internet households April 2008[a]	Complete panel April 2008	χ^2 test: Internet vs. complete panel	Dutch population 2008	χ^2 test: complete panel vs. population
One-person household	21.2%	23.7%	18.0**	35.5%	303.5**
Adults aged 65 and over[b]	9.4%	11.4%	42.5**	14.8%	148.6**
Children in the household	45.5%	42.9%	13.7**	34.8%	142.8**
Household composition:					
Single	21.7%	24.1%	23.8**	35.5%	367.5**
Couple without children	32.4%	32.5%		29.0%	
Couple with children	39.8%	37.3%		28.4%	
Single with children	5.4%	5.4%		6.4%	
Other	0.6%	0.7%		0.7%	
Regional urbanization level[b]:					
Very high	13.5%	13.6%	0.4	19.4%	439.7**
High	25.9%	25.8%		23.1%	
Moderate	22.3%	22.2%		18.3%	
Low	22.9%	22.9%		19.4%	
Without urbanization	15.5%	15.6%		19.8%	
Home owner	73.6%	70.9%	18.0**	58.9%	296.7**
Household size					
1	21.2%	23.7%	22.2**	35.5%	365.2**
2	35.7%	35.9%		32.7%	
3	14.3%	13.5%		12.4%	
4	20.3%	18.9%		13.4%	
≥5	8.5%	8.0%		5.9%	
Origin[b]					
Western Immigrant	7.1%	7.0%	0.5	9.4%	81.7**
Non-Western immigrant	6.7%	7.0%		9.5%	
Autochtonous	86.2%	86.0%		81.1%	
Voting behavior most recent elections for the parliament (2006)					
Voted at elections	85.0%	85.5%	1.30	80.4%	49.8**
Voted christen democrats (CDA)	24.2%	24.8%	0.9	29.3%	45.6**
Voted socialist party (SP)	15.4%	15.8%	0.4	16.7%	15.8

[a]In April 2008 the panel was complete, with the main recruitment campaign of 2007 completed, including all non-response follow-up efforts. [b]Denoted at individual level since population data are not available at household level.
*$p < .05$, **$p < .01$

Internet survey, and to market research online panels. When comparing the unweighted data, the LISS panel was close to the traditional survey and better than the Internet survey and online panels on five of the seven characteristics tested. The exceptions were found in the oldest age group, over 70 years and in the non-Internet households. The elderly, non-Internet group was best represented in the traditional, face-to-face survey but far better covered in the LISS panel than in the other online panels in the study.

In Table 4 we show whether the inclusion of the non-Internet households contributes to the panel representativeness. Included are the same demographic characteristics that were evaluated in the studies of Knoef and De Vos (2009); Van der Laan (2009), and Scherpenzeel and Bethlehem (2011) as well as the same variables of voting behavior included in Scherpenzeel and Bethlehem (2011). The column on the far right of the table lists the figures for the entire Dutch population (see www.statline.nl). We consider households of which at least one household member is active in the panel. For the age and urbanization level variables, population data are available only for individuals, hence we include all members of the active panel households.

If we examine the composition of the panel in April 2008 with and without the non-Internet households, then we notice that the complete LISS panel, with both Internet and non-Internet households, more closely approximates the composition of the overall Dutch population for all variables, than a panel consisting of only Internet households. Thus, 9.3% of the Internet households are over 65 years; for the complete panel this figure is 11.2%, which is closer to the overall population percentage of 14.8%. The complete panel also contains more one-person households (23.7%) and less households with a self-owned dwelling (70.9%) than a panel consisting of only Internet households (respectively 21.2% and 73.6%), so that the complete panel more closely resembles the overall Dutch population. The percentages shift for the various variables with 2 to 3 percent after inclusion of the non-Internet households, except for urbanization levels for which the difference is smaller.

Table 5
Comparison of the LISS Panel With and Without Non-Internet Households

Variable	Panel with only Internet households April 2008	Complete panel April 2008	Online vs. complete panel
Highest education in the household			
Primary Education	3.9%	5.4%	
VMBO	16.9%	18.5%	
HAVO/VWO	9.0%	8.7%	$\chi^2 = 42.4$**
MBO	26.8%	25.9%	
HBO	30.1%	28.9%	
University	13.3%	12.6%	
Net household income			
Up till 1150 Euros	6.2%	7.6%	
1151–1800 Euros	16.6%	18.4%	
1801–2600 Euros	22.7%	22.3%	$\chi^2 = 33.1$**
More than 2600 Euros	46.0%	43.4%	
Unknown	8.6%	8.4%	
Survey attitude (avg. of nine items, 7-point scale, individual level)	5.01	5.37	$t = 8.97$**
Life satisfaction (scale: 0–10)	7.45	7.58	$t = 1.83$

*$p < .05$, **$p < .01$

Table 5 contains another comparison between a LISS panel with and without non-Internet households on variables, for which no population data are available. The analysis reveals that the education level of the complete panel is significantly lower than for the panel with only Internet households. The percentage of households with only primary education is higher (5.4% vs. 3.9%) and the percentage of households with a university degree is lower (12.6% vs. 13.3%). In a similar vein the average income of the complete panel is lower than for panel with only Internet households. As expected non-Internet households are somewhat lower educated and have lower incomes than Internet households. Furthermore we find that the survey attitude of the complete panel is higher than for the panel with only Internet households.[5] We do not find differences in life satisfaction between Internet households and non-Internet households ($p > .05$). To sum up, we find that the inclusion of non-Internet households tends to bring some characteristics of the panel significantly nearer to the Dutch population. Nevertheless, significant and substantial differences between the panel and the Dutch population remain.

[5]Average of nine items, 7-point scale, items can be found at www.lissdata.nl, Core study Personality.

Step 4: Do households without Internet change after joining an Internet panel?

A potential problem as a result of providing Internet access to non-Internet households is that they may start to become more like Internet households after joining the panel. After all, the SimPC allows them to start using the Internet for purposes other than completing questionnaires. We refer to this as Internet-adoption or Internet use by the non-Internet households, as a result of panel participation. The Internet use of non-Internet households may subsequently change other aspects of their life, such as leisure spending, news gathering or social contacts.

To examine Internet use by panel members, we consider all persons that completed the questionnaires about Internet use in both 2009 and 2010. In the core questionnaire 'Social integration and leisure time', which is administered every year to all panel members in the LISS panel, we ask the panel members whether they ever use the Internet, besides for the purpose of completing questionnaires. If the respondent replies affirmatively, we then ask how many hours a week he or she spends on the Internet (including for email), in addition to completing LISS panel questionnaires. We consider a panel member as Internet user if he uses the Internet for other purposes than completing questionnaires at least one hour per week.

We examine how many people used the Internet for more than one hour a week, for purposes other than completing questionnaires. Table 6 shows that, in 2010, more than half of the panel members in non-Internet households (58.7%) starting using the Internet for at least one hour a week. However, we do not find that this Internet adoption increases over time, given how the percentage for the same group is only marginally higher in 2010 compared to 2009 (56.5%). By contrast, panel members in Internet households are more frequent users than the members of non-Internet households, the usage rate differs almost 30%-points (86.4%), χ^2 (1, N = 4730) = 273.1, $p < .001$. Remarkably, part of the panel members living in Internet households also uses the Internet for no other purpose than to complete panel questionnaires. In 2010 this concerned 13.6% of the members of Internet households, which is a slight drop compared to 2009 (14.1%). This resembles with findings of Kraut et al. (1999) who found that though individual household members affect each other's Internet usage, they showed individual usage patterns that can strongly vary within households.

In 2010, users in Internet households spent an average amount of 13.04 hours a week on the Internet. In non-Internet households this was 3.82 hours less, namely 9.20 hours, t (4728)= −4.95; $p < .001$.

Table 6
Percentage of Panel Members That Use the Internet for at Least One Hour a Week and the Number of Hours Spent on the Internet (Aside From Completing Panel Questionnaires), in Internet Households and Non-Internet Households

	Internet households[a]	Non-Internet households[a]	Sig.	Total
Use in January 2009	85.9%	56.5%	**	82.8%
Use in January 2010	86.4%	58.7%	**	83.5%
# hours a week 2010 (users)	13.04	9.20	**	12.79
N	4'238	492		5'558

[a]Panel members that completed questions about Internet use in 2009 and 2010.
*$p < .05$, **$p < .01$

The questions about Internet use were followed by a conditional question about the type of Internet activities engaged in. The question read: Can you indicate whether you ever engage in the following online activities? A list of different activities was then displayed.

We examine the activities by panel members that use the Internet for at least one hour a week. Table 7 shows that panel members in Internet households use all Internet applications more frequently than non-Internet household members (see Table 7), with the exception of reading news and magazines, χ^2 (1, N = 4730) = 0.2, p = .65. In both groups, over half of the panel members that use Internet read news and magazines (resp. 53.2% and 52.0%). Most persons active on the Internet in both Internet and non-Internet households use email (resp. 98.2% and 91.5%) and use the Internet to search for information (resp. 97.0% and 88.6%). However, whereas the majority of Internet users in Internet households engage in online banking (83.8%), this applies to only half of the Internet users in non-Internet households (52.0%). Substantial differences between the two groups also emerge for purchasing products and downloading music and films. Thus, 67.9% of the individuals in Internet household members sometimes purchase products online, while just 34.9% of the individuals in non-Internet households do so. Almost one-third of Internet household members download music and films, against just 12.2% in non-Internet households. Chatting using e.g., MSN is an activity for 29.4% of the Internet household members, compared to 18.8% of users in non-Internet households. In sum, the individuals in non-Internet households who start to use the

Internet for other purposes than completing questionnaires use less of the options offered by Internet, and mainly use the simpler applications compared to persons living in Internet households.

Table 7
Use of Internet Applications by Panel Members in Internet Households versus Non-Internet Households (Internet Users Only)

Variable	Internet households	Non-Internet households	χ^2 test
Email	98.2%	91.5%	66.2**
Searching information	97.0%	88.6%	66.9**
Online banking	83.8%	52.0%	202.9**
Purchasing products	67.9%	34.9%	158.6**
Reading news/magazine	53.2%	52.0%	0.2
Downloading music/films	31.1%	12.2%	55.8**
Chat/ MSN	29.4%	18.8%	18.3**
N	5'015	352	

Note. Individuals that answered questions about Internet use in 2010.
*$p < .05$, **$p < .01$

Conclusion and discussion

This article examines whether providing non-Internet households with Internet facilities, which requires a substantial investment, significantly improves the quality of an Internet panel. The small segment of households without Internet access in the Netherlands differs from the rest of the population with regard to household size, age, origin and place of residence. This underlines the importance of including them in Internet panels.

At the start of 2010, the LISS panel contained 545 households that did not have Internet access at the time of recruitment, constituting approximately 10% of the entire panel. These participants formed a representative sample of non-Internet households in terms of ethnical background, region, and household composition, but not in terms of age: higher age groups without Internet are somewhat underrepresented in the panel. Furthermore, it appears that approximately 15% of the panel members from Internet households use the Internet only for the purpose of completing questionnaires for the panel. Both groups are probably hardly covered by regular Internet panels.

In addition, we find that non-Internet households that are part of the LISS panel are more loyal than recruited Internet households. Non-Internet households show lower attrition rates and higher response rates to the individual questionnaires than participants from Internet households. Furthermore, older Internet households show an above-average attrition rate, which is not the case with older non-Internet households. It is quite possible that the households without Internet access feel a certain sense of reciprocity or obligation towards the panel organization for the facilities that they have received. Besides, they would lose these facilities again if they were to leave the panel.

If we examine the composition of the panel in April 2008 with and without the non-Internet households, then we notice that the complete LISS panel, with both Internet and non-Internet households, more closely approximates the composition of the overall Dutch population for all demographic variables we tested, than a panel consisting of only Internet households. The inclusion of non-Internet households thus contributes to the representativeness of the panel, as was also shown by Hoogendoorn and Daalmans (2009). Nevertheless, some significant and substantial differences between the LISS panel and the population statistics remain, even when non-Internet households are included. However, as Scherpenzeel and Bethlehem (2011) have shown, on many demographic characteristics and on voting behavior the LISS panel comes close to a traditional face-to-face survey and is better than an Internet survey based on volunteer sampling and than online panels excluding non-Internet households.

The question is whether the benefits of including non-Internet households are higher than the costs (in terms of costs of the computers and time efforts of support and help). The answer to this question depends on the purpose of the study and the Internet penetration of the population. For studies, either scientific or commercial in nature, which require very precise measurements of specific concepts within a population (e.g., electoral polls, employment measurements, income) the benefits of improving the representativeness of the panel can be substantial, whereas for causal or small-scale experimental studies the benefits are less straightforward. Also, studies that investigate topics that relate to Internet use in society, e.g., smartphone penetration or social media use, can yield strongly biased results if non-Internet households are ignored (Leenheer & Elsen, 2012). Furthermore, in

countries with low Internet penetration, inclusion of non-Internet households may change the research results substantially and may provide competitive advantage even in the commercial sector. It is especially in these countries and regions that including non-Internet households in an Internet panel can pay off.

Approximately half of the panel members from households without Internet access (58.7%) start to use the Internet after joining the panel. Literature on innovation adoption shows that the opportunity to try out an innovation contributes to the acceptance of the innovation. However, as Mahajan, Muller, and Srivasta (1990) have shown, people belonging to the group of "late adopters" of computers make much more limited use of computers after becoming users. Something similar seems to apply here to Internet use. More complex and risky applications such as online banking or downloading music are used much less frequently by non-Internet households, who use it mostly for the purposes of email and searching for information. For this reason one may question whether they have indeed truly adopted the Internet, or whether they are still in the early stages of the adoption process. This question may perhaps be answered by monitoring the behavior of LISS panel members that did not previously have Internet access over the coming years. Analyses of data from 2009 and 2010 show only minor shifts. Furthermore research of Anderson and Tracey (2001) shows that household's Internet adoption does hardly change other aspect of their life, suggesting that the use of Internet by some of the non-Internet households will not jeopardize panel representativeness

Implications for further research

Internet panels describe a steadily growing segment of the population, but still fail to reach some parts of it. In 2011 in the Netherlands, about 9% of households do not have Internet access; in Europe as a whole, this number is over 40%. In socio-demographic terms this mainly concerns older households, single-person households and households with members of non-Western origin. In different academic areas (health care studies, leisure studies, public policy) senior citizens and multicultural groups are increasingly attended to, as they represent an increasingly important population segment. These groups are not only underrepresented in many online panels, but it may also well be that participants from these groups are not fully representative of these segments. Senior citizens aged over 80 who are active on the Internet and actively register for Internet panels are perhaps more innovative and progressive with regard to other lifestyle characteristics (e.g., use of tablets or smartphones) than this age group tends to be (Eastman & Iyer, 2005).

Most Internet panels, and especially those for which people can register at their own initiative, do not cover the entire population in terms of certain socio-demographic factors, nor with regard to psychographic characteristics and other features that are more difficult to observe. It may thus be expected that households with low technological knowledge and interest do not have an Internet connection and are therefore not represented in Internet panels. Caution is especially required when using Internet panels to perform research into innovations and technology. The segment that is not interested in such innovations as the smartphone, the I-pad or mobile banking may well be larger than the results of some Internet panels may show.[6]

Another implication lies in international survey research, in which research results from various countries are compared. As said before, the differences in Internet use between countries are great, even just among Western countries. An Internet panel in a country with high Internet use represents a different section of the population than an Internet panel in a country with low penetration. In such cases, results obtained from online research in different countries cannot be compared.

Providing Internet access to households that do not yet have such access seems a rather drastic means of recruiting non-Internet households, and at present still requires a substantial investment. However, with the advent of smart phones and the Mobile Web, faster and more affordable solutions are becoming more easily available. A good example is the previously mentioned new Internet panel ELLIPS in France, which provides tablets, a much cheaper option than computers, to the respondents who do not have Internet access. The coverage bias may decrease quickly given the rapid pace of mobile web diffusion throughout European countries (Fuchs & Busse, 2009). Importantly, our results show that part of the members of Internet households also do not use the Internet. These persons are represented in the LISS panel, recruitment for which takes place at household level using traditional methods, but are most likely missing in many other Internet panels. This group of potential panel members does not need to be provided with Internet facilities, but they do need to be approached through other channels than the Internet and may need some help getting started. Future studies could look in more detail into the participation patterns and Internet activities of these panel members.

[6]Additional analyses within the LISS panel show that in 2012 only 3.2% of the individuals in non-internet households access the internet with a smartphone and 0.5% with a tablet.

This paper shows that including non-Internet households in an Internet panel can successfully be done by providing them necessary appliances; the non-Internet households are different from Internet households and remain so after having been members. A limitation, however, is that our research studies a single panel in a country with very high Internet penetration. On the one hand, it may be considered a strong point that even in a country with high Internet penetration including non-Internet households improves the panel significantly. On the other hand, similar panels have been build in the past, when Internet penetration was still much lower, and in other countries having a lower Internet penetration rate. Examples of the former are the CentERpanel, operated at CentERdata as well, and the KnowledgePanel in the U.S., operated by Knowledge Networks. A recent example of the latter is the French panel ELLIPS that was mentioned earlier. The Internet penetration in France is lower than in the Netherlands (in 2012 it is estimated to be 77%, www.Internetworldstats.com) and the coverage of the Internet providers is not country-wide. Therefore, the panel provides tablets, using the mobile networks, to the respondents who do not have Internet access. This effort illustrates that with the newest technology, the inclusion of non-Internet households in a panel is also possible in much larger countries with a lower Internet penetration than the Netherlands. Nevertheless, we do not know whether the procedure used to include non-Internet households in an online panel can be copied in a similar way to countries with very low penetration levels. Neither do we have statistical proof of how it would improve Internet panels in low Internet penetration environments. Further research is needed to address these issues.

Acknowledgements

We thank Mick Couper, Don Dillman and participants of the AAPOR Conference and MESS-Workshop for valuable input.

References

Anderson, B., & K. Tracey (2001). Digital Living: The impact (or otherwise) of the Internet on everyday life. *American Behavioral Scientist*, *45*(3), 456–475.

Brown, S. A., & Venkatesh, V. (2005). Model of adoption of technology in households: a baseline model test and extension incorporating household life cycle. *MIS Quarterly*, *29*(3), 399–426.

Callegaro, M., & Disogra, C. (2009). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*(5), 1008–1032.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*, 464–494.

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, *36*, 131-148. doi: 10.1016/j.ssresearch.2005.10.002

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR)) and the Internet. *Social Science Research*, *38*, 1–18.

Dillman, D. A., Reips, U., & Matzat, U. (2010). Advice in surveying the general public over the Internet (Editorial). *International Journal of Internet Science*, *5*, 1–4.

DiSogra, C., Callegaro, M., & Hendarwan, E. (2009). Recruiting Probability-Based Web Panel Members Using an Address-Based Sample Frame: Results from a Pilot Study Conducted by Knowledge Networks. Presented at the Joint Statistical Meetings.

Eastman, J. K., & Iyer, R. (2005). The impact of cognitive age on Internet use of the elderly: an introduction to the public policy implications. *International Journal of Consumer Studies*, *29*(2), 125-136. doi:10.1111 /j.1470-6431.2004.00424

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

Fuchs, M., & Busse, B. (2009). The coverage bias of mobile web surveys across European countries. *International Journal of Internet Science*, *4*(1), 21–33.

Hoogendoorn, A. W., & Daalmans, J. (2009). Nonresponse in the recruitment of an Internet panel based on probability sampling. *Survey Research Methods*, *3*(2), 59–72.

Knoef, M., & de Vos, K. (2009). The representativeness of LISS, an online probability panel. Working paper. CentERdata, Tilburg University, Available at: http://www.lissdata.nl/lissdata/About_the_Panel/Composition_& _Response

Kraut, R., Mukhopadhyay, T., Szczypula, J., Kiesler, S., & Scherlis, B. (1999). Information and communication: alternative use of the Internet in households. *Information Systems Research*. *10*(4), 287-303.

Leenheer, J., & Elsen, M. (2012). Studying the Internet through the Internet: An assessment of mobile Internet and social media use in the Netherlands using the LISS online panel. Working paper. CentERdata, Tilburg University

Loosveldt, G., & Sonck, N. (2008). An evaluation of the weighting procedures for an online access panel survey. *Survey Research Methods*, *2*(2), 93–105.

Mahajan, V., Muller, E., & Srivastava, R. K. (1990). Determination of adopter categories by using innovation diffusion models. *Journal of Marketing Research*. *27* (1), 37–50.

Rookey, B. D., Hanway, S., & Dillman, D. A. (2008). Does a probability-based household panel benefit from assignment to postal response as an alternative to Internet-only? *Public Opinion Quarterly*, *72*(5), 1–23. doi: 10.1093/poq/nfn061

Scherpenzeel, A. C., & Betlehem, J. (2011). How representative are online panels? Problems of coverage and selection and possible solutions (M. Das, P. Ester, L. Kaczmirek, Eds). Chapter 5. In *Social research and the Internet: Advances in applied methods and new research strategies* (pp.105-132). *Boca Raton: Taylor & Francis.*

Scherpenzeel, A. C. (2010). LISS Panel Statistics 2010. Available at: http://www.lissdata.nl/lissdata /About_the_Panel/ Composition_&_Response

Scherpenzeel, A., & Zandvliet, R. (2011). Sleepers and inactives in online panels (Slapers en inactieven binnen online panels). In: Ontwikkelingen in het marktonderzoek: Jaarboek MarktOnderzoekAssociatie, 36, 189–204. Spaar[en] Hout Haarlem.

Scherpenzeel, A., & Toepoel, V. (2012). Recruiting a probability sample for an online panel: Effects of contact mode, incentives and information. Public Opinion Quarterly, 76 (3): 470–490.

Scherpenzeel, A. (Forthcoming). Survey participation in a probability-based Internet panel in the Netherlands. In: U. Engel, B. Jann, P. Lynn , A. Scherpenzeel, P. Sturgis (Eds.). Improving survey methods. England: Taylor & Francis.

Van der Laan, J. (2009). Representativity of the LISS panel. Discussion Paper 09041. The Hague/ Heerlen, The Netherlands: Statistics Netherlands.

Zhang, C., Callegaro, M., & Thomas, M. (2008). More than the Digital Divide? Investigating the differences between Internet and non-Internet users on attitudes and behaviors. Working paper, MAPOR conference, Nov. 21–22.