Survey Research Methods (2008)
Vol.2 , No.2 , pp. 93-105
ISSN 1864-3361
http://www.surveymethods.org

An evaluation of the weighting procedures for an online access panel survey

Geert Loosveldt
University of Leuven, Belgium

Nathalie Sonck
University of Leuven, Belgium

Opinion research is frequently carried out through the Internet and a further increase can be expected. The article focuses on the online access panel, in which respondents are previously recruited through non-probability methods. Despite substantial time- and cost-reduction, online access panel research mainly has to cope with limited Internet coverage and self-selection in the recruitment phase of new panel members. The article investigates whether frequently applied weighting procedures, based on poststratification variables and propensity scores, make online access panel data more representative of the general population. To address this issue, the answers to identical questions are compared between an online self-administered survey of previously recruited online access panel respondents and a face-to-face survey of randomly sampled respondents of the general population. Both respondent groups were surveyed at a similar moment in time (2006-2007) in the same geographical region (Flanders, Belgium). The findings reveal many significant differences, regarding sociodemographic characteristics as well as attitudes towards work, politics and immigrants. The results can be explained by both the specific characteristics of the respondent groups and mode effects. Weighting adjustment had only a minor impact on the results and did not eliminate the differences.
Keywords: online access panel, representative samples, weighting procedures, propensity score adjustment

Introduction

Opinion research is frequently carried out through the Internet and a further increase in online surveys can be expected, because of continuous technological progress, the rising penetration of the Internet in everyday life, and a massive increase in web usage (Vehovar et al. 2002:229). A specific type of web surveys which is increasingly used to measure peoples opinions is the online access panel survey. An important question regarding this relatively new survey technique is how representative the obtained results are of the general population. It can be expected that such an online access panel is composed of people familiar with the Internet and willing to regularly spend some time filling out online surveys. It is important to study the degree to which this specific group of online access panel respondents differs from the general population in sociodemographic terms, as well as with regard to attitudes about various societal themes.

Couper (2000:482-484) argues that "it is most problematic when surveys based on volunteer panels of Internet users claim they are representative of a more general public". Bias in online access panel surveys is primarily caused by limited Internet coverage and self-selection in the recruitment phase of new panel members. In order to overcome these problems, weighting procedures based on post-stratification variables and/or propensity scores are frequently applied to the online access panel sample. It can be questioned whether these frequently used weighting adjustments make online access panel data more comparable with traditionally obtained survey results. To address this issue, this study will compare the answers to identical questions between an online self-administered survey of previously recruited online access panel respondents and a face-to-face survey of randomly sampled respondents of the general population.

Contact information: Geert Loosveldt, Centre for Sociological Research (CeSO), Katholieke Universiteit Leuven, Parkstraat 45, Box 3601, B-3000 Leuven, Belgium, E-mail: Geert.Loosveldt@soc.kuleuven.be

Background

Following Coupers typology of web surveys (2000:477), several types of Internet panels can be distinguished based on their methods of selecting new panel members to become regular respondents. This article only focuses on the online access panel type. In basic terms, an online access panel makes use of the web survey technique to survey previously recruited respondents who are willing to participate regularly in surveys. People are invited to join the panel through non-probability methods such as selfselection (Couper 2000:477-490). The channels through which new panel members are recruited can be online, offline or a combination of both so that low-frequency users of the Internet also have a chance of being selected. Since the adopted panel recruitment methods are directly linked to selection bias, a multi-method recruitment is suggested as the most effective (Stenbjerre and Laugesen 2005:9).

Although in a short fieldwork period and at a relatively low cost a large group of people can be reached, from whom answers can be rapidly collected and analysed due to extensive automation (Bandilla et al. 2003:235), there are also

disadvantages to online access panel surveys, of which the most important are non-coverage and sample selection difficulties (Bandilla et al. 2003; Dillman 2000; Couper 2000). These methodological problems are considered next, since they may have important implications for the representativeness of the survey data for the general population. The online access panel sample will eventually consist of only those people who could be reached through the Internet (coverage), joined the online access panel (selectivity) and responded to the survey invitation ((non-)response). Since a possible solution for reducing these biases in online access panel surveys is weighting (Taylor 2005; Bethlehem and Stoop 2007), this is discussed subsequently.

(Non)coverage Coverage error occurs when not all elements of the survey population have a known non-zero chance of being sampled for survey participation. The main reasons why the Internet can not be used as a proper sampling frame is because it does not include all members of the general population, and those people who can be included differ from those who cannot. Possible bias is thus not only related to the number of people who have access to the Internet but also to the average differences between the persons with and without Internet access (Dillman 2000:354-356). Although more people have gained access to the Internet over time, important socio-demographic differences still exist. In the US, for example, Internet users are more likely to be young, male, white, more educated, wealthy, city residents and the parents of children living at home (Lenhart et al. 2003; Vehovar et al. 1999:966; Taylor 2005:2).

Selectivity of online access panels Aside from the non-coverage problem, the main difficulty is implementing a probability-based sample survey on the web (Couper 2000:476). By adopting nonprobability, and in particular self-selective panel recruitment methods, the chances of being sampled are unknown. For online access panel research, self-selection is mainly a problem during the recruitment of new panel respondents, rather than at the point of deciding whether they will actually respond to the online survey invitation. It may be assumed that the people who participate in volunteer online access panels will have specific characteristics; they will have enough time available and have the skills required to regularly participate in surveys on the Internet. As a result, their answers to attitude questions could differ substantially from randomly selected persons in the general population. To study this selectivity bias, a comparison of the survey results between an online access panel and a representative sample is necessary.

For example, Bandilla et al. (2003:236) compared the results of a web survey collected from a pre-recruited panel of Internet users in Germany with those from a self-administered mail survey collected among the general population. After adjusting the sample of Internet users for basic socio-demographic characteristics, significant differences were still found between the two survey modes. However, a higher level of agreement across both methods could be found in parallel respondent groups, mostly among the more-educated people (Bandilla et al. 2003:238-241). Vehovar et al. (1999) observed significant differences in four of the seven items concerning electronic trade between a self-selective sample of web respondents and a telephone survey of frequent Internet users; the differences were largely due to the differences in Internet experience between the samples. Additionally, important social differences seem to exist between Internet users and non-users. Research showed that people with greater social trust and greater subjective control over their lives tend to use the Internet more frequently (Lenhart et al. 2003). Moreover, through practical survey experience it was found that online panels comprise more politically- and socially-active respondents (Duffy et al. 2005:627).

Possible solution: weighting?

A possible solution to reduce the biases due to limited coverage and self-selection may be weighting adjustment (Bethlehem and Stoop 2007:123). Taylor (2005) argues that both post-stratification and propensity weighting procedures are required in order to reduce the biases in raw online access panel results (Taylor 2005:3). The poststratification weighting aims to adjust for demographic differences between the sample and the population under consideration (Duffy et al. 2005:620). However, the problem of selection bias cannot always be solved with the application of such post-stratification procedures since some variables of interest do not show a sufficiently strong relationship with the demographic weighting variables (Bethlehem and Stoop 2007:123). Moreover, post-stratification weights can correct the online sample for proportionality but not necessarily for representativeness. Adjusting the proportional over- and underrepresentation of certain respondent groups does not mean that the substantive answers of online access panel respondents also become comparable to those of the general population.

It is precisely because attitudinal and behavioural differences are still observed even after applying post-stratification weights based on demographic characteristics, that another weighting technique which has been primarily used in the context of experimental designs (e.g. Rosenbaum and Rubin 1983) is applied to online panel samples, namely the propensity score adjustment. This weighting technique aims to correct for differences caused by the varying inclination of individuals to participate in online access panel surveys (Duffy et al. 2005:620). To this end, a probability-based reference survey that is not conducted over the Internet is used. It is assumed that this reference survey produces unbiased estimates (Bethlehem and Stoop 2007:123). Through logistic regression, the probability of each respondent participating in the online access panel survey is estimated, based on background characteristics.[1] People are then classified into groups with a similar propensity score so that an equal distribution is obtained of the regression-used characteristics,

[1] Aside from demographic variables, 'webographic' or lifestyle variables may be used (cf. Schonlau et al. 2007).

both in the online panel and another, more traditional survey sample (Lee 2006; Schonlau et al. 2006).

Although the post-stratification and propensity-based weights are necessary in order to compare online survey results with survey results from the general population (Duffy et al. 2005:620), their application in research practice produces rather diverse results (Vehovar et al. 1999:966). Malhotra and Krosnick (2007), for example, found that weighting adjustments could not eliminate the significant differences observed between two face-to-face surveys and two online access panel surveys regarding demographic and political variables. In some cases, these differences became even larger after weighting (Malhotra and Krosnick 2007:293-296). Taylor (2005:5) observed that some differences between online and telephone survey data disappeared completely when using post-stratification weighting, some only after propensity score adjustments, and others continued to exist after applying both weighting procedures. Duffy et al. (2005) reached a similar conclusion, noting that certain questions produced similar results for online and face-to-face surveys without weighting, that other questions were comparable only after weighting (e.g. voting intentions), and that yet other topics resulted in significant differences after the application of both post-stratification weighting and propensity score adjustments (e.g. political activism).

In order to evaluate these weighting procedures that are frequently applied to online access panel data, it is important to have a proper reference survey to compare them with. Since the main concern of this article is the representativeness of online access panel data compared with survey results obtained from the general population, a randomly selected sample is required. In comparison with current mail or telephone surveys, the face-to-face survey method typically has higher response rates and greater potential to represent the general public (Malhotra and Krosnick 2007:287). Therefore, a probability-selected sample of personally-interviewed respondents is considered to be the most appropriate for this bench-mark study. However, by comparing a face-to-face interview with an online panel sample a new problem is inevitably introduced, namely mode effects. Although the focus of this article is an evaluation of weighting adjustments to an online access panel sample in order to obtain more representative data, the possibility of mode effects influencing the results of the comparison between the different survey methods cannot be ignored and is therefore considered in the next Section.

A confounding issue: mode effects

Mode effects or the existence of systematic differences between data collected by a face-to-face interview and an online survey can be expected a priori, based on three main classes of factors which are highly interrelated: media related factors, information transmission differences and interviewer effects. The media related differences between modes refer to the social conventions and customs associated with the communication medium under consideration, such as the degree of familiarity with the mode and who controls the survey situation (the so-called 'locus of control'). The control over the pace and flow of a face-to-face interview lies primarily with the interviewer, but it is the respondent who chooses when, where and at what pace online questions are answered (de Leeuw 1992:26-29; de Leeuw 2005:244).

The second group of factors concerns the information transmission differences between modes that arise depending on which communication channels can be used, namely verbal, nonverbal and/or paralinguistic (e.g. emotional tone, timing and emphasis) communication (de Leeuw 1992:29-30). A related classification is based on the presentation of the stimuli; visually or auditory (Schwarz et al. 1991:194-195; Dillman 2000). A face-to-face survey can make use of all three communication channels and both types of stimuli presentations. On the other hand, the online survey is primarily presented in a visual way and without nonverbal communication. This requires more demanding cognitive processing from the online respondents and may potentially result in lower data quality. Recently, Heerwegh and Loosveldt (forthcoming) found evidence that online respondents provide more 'dont know' responses, more non-differentiation on rating scales, and a higher item non-response rate compared with face-to-face respondents. This is referred to as a 'satisficing' effect, which can arise depending on the respondent's ability and motivation as well as the difficulty of the survey task (Holbrook et al. 2003:82).

The third class of mode effects concerns the impact of the interviewer on the question-answer process. A potential positive impact may occur because the interviewer can motivate the respondents, clarify questions and give additional information when needed. However, a negative impact of the presence of an interviewer is also possible, because it may reduce people's feelings of privacy and anonymity (de Leeuw 1992:31-32). This may in turn affect people's willingness to provide sensitive information and thus the degree of social desirability (de Leeuw 2005:245). The assumption is that the greater the "social distance" between respondent and interviewer, the greater the tendency is to answer honestly. In a face-to-face survey situation, the social distance is assumed to be smaller than in an online mode, which makes people more susceptible to possible disapproval signs from the interviewer (Holbrook et al. 2003:86-87). For example, Comley (2003:7-8) found that online marketing surveys produce a higher degree of 'honest' answers than interviewer-administered surveys. With regard to political questions, Duffy et al. (2005:624-625) found that in the absence of an interviewer, online respondents are more likely to display their voting intention than personally-interviewed people.

Data and method

In order to study the weighting adjustments that are frequently applied to online panels, an empirical comparison is performed between the data obtained from a self-administered online access panel survey and the answers collected in a face-to-face interview with a probability-based sample of the general population.

Databases and questionnaire

The Belgian dataset of the third round of the European Social Survey (ESS 2006) is used. The ESS aims to study changing public attitudes and values in Europe, as well as to improve methods of cross-national survey measurement. It is an academically-driven face-to-face survey, organized every two years in over 20 European countries with a high-standard survey protocol (including refusal conversion). In Belgium, a two-stage probability sampling design is implemented, stratified according to provinces and municipalities in which persons aged 15 years and older are selected by simple random sampling. No incentives were provided. The response rate for the Flemish region was 62.25 percent (n=980).

The online access panel in this study is composed of (some 80,000) panel members who are recruited through different channels, both online and offline. The panel institute (iVOX, a Belgian online market research company) does not use probability-based methods to recruit new panel members, but potential respondents are invited to join the panel through pop-ups and other web advertisements (online), as well as through traditional survey methods in follow-up surveys (offline). Upon panel registration, people are asked to complete a short online questionnaire about their personal background. Based on the information database, a stratified sample is drawn of adult panel members aged between 18 and 74 years old, living in the Flemish region of Belgium. The sample is stratified according to gender, age and education, based on census information. Additionally, certain groups of respondents are over- or undersampled according to previously achieved response rates within the stratified cells. As an incentive, the respondents could win gift vouchers from multimedia shops. From the 5685 survey invitations sent by e-mail to the online panel members, 3235 online surveys were completed.[2]

It is relevant to mention the substantial difference in rapidity between both modes through which the survey data is collected. The responses of the face-to-face interviews are gathered over almost four months. The online survey on the other hand is accessible for the selected panel members to be answered during only one month. The same day the e-mail invitations were sent to the selected panel members, already 1576 respondents completed the web survey. This means that 28 percent of the survey invitations sent were responded to on the very first day of data collection. In other words, almost half of the completed surveys (48.7 %) were filled out during the first fieldwork day. After 10 days a first, and after another 10 days a second, reminder was sent by e-mail to the panel members who had not yet responded to the survey invitation. After the second reminder, there was only a response supplement of 4.5 percent of the sent invitations. Largely due to practical reasons, it is almost impossible to obtain a similar kind of rapidity in the collection of data from personal interviews for which appointments have to be made and interviewers have to visit the respondents. In addition to time, there is an important difference in cost between both surveys. The cost per completed interview for the face-to-face survey was about €140. For the online panel survey there is a fixed cost of the panel management and the variable cost of about €3.

With regard to the questionnaire, several identical questions regarding different social and political themes were asked in both surveys. In the online survey, however, a shorter version of the questionnaire was used. In order to reduce the differences in the questionnaires between the two survey modes, the same question wordings and answer categories were used. In the online survey, all the answer categories were visually provided on the computer screen. In the face-to-face survey, all the response categories were orally presented to the respondent, and for most of the attitude questions, the response options were also visually shown on response cards.

In order to adequately compare the results of both surveys, only the respondents from the Flemish region aged between 18 and 74 years old have been taken into consideration. In this region, an increase in the percentage of inhabitants owning an Internet connection has been observed recently, from 33 percent in 2001 to 62 percent in 2006 (SCV survey 2006). With regard to Internet usage (% of people who used the Internet in the last three months), Figure 1 shows significant differences for Flanders when studied against some basic characteristics. It can be seen that Internet usage declines with age. Furthermore there are slightly more men than women who use the Internet. People who completed more years of education and people who have paid jobs also use the Internet slightly more. These results illustrate the typical coverage problem of online surveys and therefore the necessity of weighting the results of online samples.

Weighting procedures

An initial evaluation of the representativeness, which serves as the basis for poststratification weighting is performed by comparing the effective samples with the target population of Flemish people between 18 and 74 years old. Table 1 presents the unweighted univariate distributions of gender, age and the highest obtained educational level for the online access panel sample, for the probability-based face-to-face sample and the target population. The population figures come from the Belgian institute of statistics, more specifically from the National Survey of Labour Force (LFS). Discrepancies between the effective samples and the population figures are tested at the global level of the variable distribution by a Chi-square test based on the expected and observed cell frequencies, as well as at the level of the individual categories by calculating 95 percent confidence intervals around the sample percentages. Since the online panel sample is

[2] This resulted in a 'completion rate' of 59.96 percent which does not take into account the self-selection at the point of inviting new panel members. Furthermore, artificially high completion rates may be obtained by eliminating groups from the panel that are difficult to reach or by maintaining those panel members who respond well (Van Ossenbruggen and Vonk 2006). Therefore, the completion rate of online access panels is not comparable to the response rate of traditional surveys.

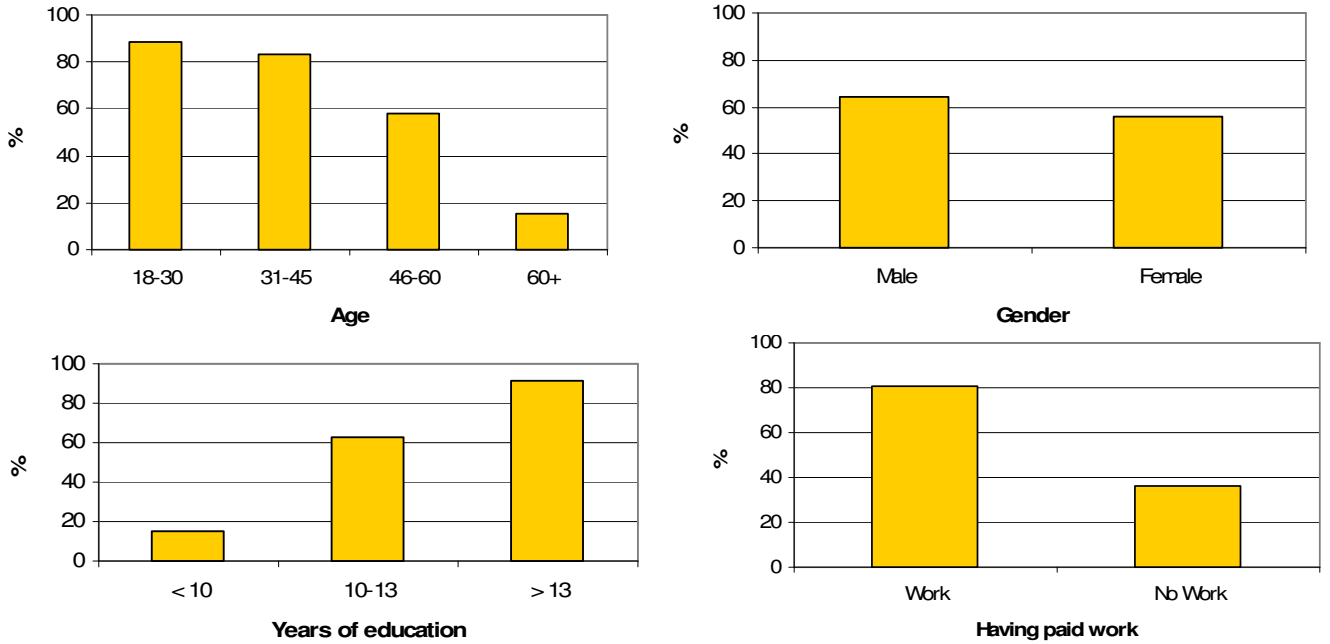

Figure 1. Percentage of Internet usage in the Flemish population by age, gender, years of education and having paid work (based on figures from SCV survey 2006)

stratified according to age, gender and education, it may be assumed that these variables are correctly represented in the effective online sample when there is no non-response bias.

Although it could be expected that more males respond to online surveys, because they are more often a (frequent) Internet user (SCV survey 2006), Table 1 shows no significant differences in gender neither for the online access panel, nor for the random sample, compared to the population figures.

With regard to age, the online panel survey significantly deviates from the distribution of the Flemish population between 18 and 74 years old. As can be expected, the oldest age group of 65-74 years is underrepresented in the web survey (8%) compared with the target population (14%). This is an indication that it is quite difficult for the online access panel to collect information from the respondents in this oldest age category. Although it could be expected that people in the youngest age category of 18-24 years old are easily reached through the Internet, they are also proportionally underrepresented in the online panel survey (7%) compared with their share in the target population (11%). As shown in Table 1, people between 55 and 65 years old are significantly overrepresented in the panel sample (25%) compared with the target population (16%). This is rather remarkable. It seems that these middle-aged respondents not only have enough time but also the skills required to participate regularly in online surveys. For the face-to-face survey of randomly sampled respondents, the global age distribution does not deviate significantly from the population (with the exception of the age category 25-34 years). This means that the randomly obtained sample which is not stratified by age gives a better representation of the different age groups compared with the stratified online panel sample.

Concerning the educational distribution of the respondents, it is the random sample that differs significantly from the population percentages. In particular, the group of respondents who have obtained (at the most) a higher secondary level are overrepresented in the random sample (58%) compared with the target population (37%). The overall distribution of education in the online panel sample does not deviate from the population figures. Based on the information in Table 1, post-stratification weights were applied to both samples. These were calculated as the ratios of the respective population and sample percentages for the combined characteristics of gender (2 classes), age (6 classes) and education (3 classes).

For the propensity score adjustment, the raw online panel data are weighted, using at least the same variables as in the post-stratification weighting. For this, a logistic regression was performed to know people's probability of participating in the online access panel. The variables of gender, age, having paid work, education and living area were included in the estimation of the propensity scores and subsequently in the calculation of the weights. Table 2 gives an overview of the logistic regression coefficients, odds ratio's and Wald tests. As can be seen from this table, all of the included variables, with the exception of gender, were found to have a significant effect on the selective participation in the online panel survey.

A rather remarkable observation is that people in their fifties (50-59 years old) are 2.16 times more likely to participate in the online panel survey in comparison with the

Table 1: Unweighted demographic characteristics of respondents compared with the population of Flanders, 18-74 years old[a]

	Sample				Target population[b]	
	Online Panel		F2F Survey		Census	
Characteristic	%	n	%	n	%	n
Gender	n.s.		n.s.			
Male	51.11	1 629	50.41	494	50.10	2 184 010
Female	48.89	1 558	49.59	486	49.90	2 175 348
Age	***		n.s.			
18-24	7.37*	237	13.06	128	11.41	497 600
25-34	17.81	573	13.67*	134	17.57	766 149
35-44	19.02*	612	22.35	219	21.43	934 037
45-54	23.22*	747	19.49	191	20.08	875 557
55-65	25.02*	805	16.53	162	15.95	695 246
65-74	7.55*	243	14.90	146	13.55	590 768
Education	n.s.		***			
Lower secondary	36.83	1 151	28.88*	283	36.39	1 586 175
Higher secondary	34.72*	1 085	57.65*	565	37.37	1 629 120
Higher education	28.45*	889	13.47*	132	29.76	1 144 062

[a]Significance levels used throughout this article: n.s. not significant at alpha level 0.05; * $p < 0.05$;** $p < 0.01$;*** $p < 0.001$

[b]Source: LFS

older age group (60-74 years old), controlling for the other variables in the regression. Furthermore, people living in a (sub)urban area are 1.97 times more likely to participate in the online panel survey, in comparison with those living in a rural area. Participation in the online panel survey is significantly lower for people who have a paid job ($\beta = -0.473$) and for people without a degree in higher education ($\beta = -0.622$ for lower secondary; $\beta = -1.325$ for higher secondary). After the propensity scores were estimated, based on the background variables included in Table 2, all the respondents were categorized into 10 groups of 10 percent. Weights were then calculated to make the distribution of the web respondents similar to that of the face-to-face respondents with regard to their propensity scores (Lee 2006; Schonlau et al. 2006).

Main findings

The survey results of the online access panel respondents were compared with the personally-interviewed respondents for different themes such as work satisfaction, political attitudes and ethnocentrism. The tables included show the observed differences in the response distributions between the two modes, their significance tested by Chi-square tests for categorical variables and two-sided t-tests for (quasi-) continuous variables. For ordinal variables, both tests were performed so that the overall mean scale values, as well as the percentage distributions, could be compared. An initial comparison is tested between the raw or unweighted data among the online panel and face-to-face survey. A second comparison is tested among both surveys after the post-stratification weighting based on the combined cells of age, gender and education. Finally, the propensity score adjusted online panel data are compared with the unweighted face-to-face survey data.

Where differences can be observed between the online and face-to-face survey data, two major explanations can be put forward. The first is a selection effect, which refers to the unique nature of the group of people who are willing to participate regularly in online access panel surveys. The hypothesis is that if the answers remain significantly different between both survey methods even after standardizing basic variables such as age, gender and education, the respondent groups are fundamentally different and therefore their survey results are not comparable. A second possibility is the existence of mode effects which occur if people give different answers depending on the method of data collection. In this case, people may answer differently depending on whether they are responding to the (non-) verbal communication of an interviewer or whether they answer the questions when sitting in front of a computer, which provides a higher feeling of anonymity and a larger "social distance" between the respondent and interviewer (Holbrook et al. 2003:86-87; Kellner 2004:16-18). As the available data do not allow for a clear distinction to be made between sample selection biases and mode effects, they will be considered concurrently in the interpretation of the results. In order to have some indication, different kinds of questions are compared for which a varying degree of social desirability can be expected. First, factual questions concerning people's living area and work situation are compared, for which the interviewer impact is deemed minor. Then, attitudinal questions which are more

Table 2: Logistic regression of the probability of participating in the online access panel (n= 4081)

Variables		β	P-value	Odds Ratio	95% Wald CL
Gender	Male	0.140	n.s.	1.151	[0.988 - 1.340]
	(Ref: Female)				
Age	18-29	0.404	**	1.498	[1.153 - 1.946]
	30-39	0.597	***	1.817	[1.367 - 2.416]
	40-49	0.501	***	1.651	[1.267 - 2.151]
	50-59	0.770	***	2.159	[1.682 - 2.770]
	(Ref: 60-74)				
Work	Paid work	-0.473	***	0.623	[0.517 - 0.751]
	(Ref: No paid work)				
Education	Lower secondary	-0.622	***	0.537	[0.421 - 0.684]
	Higher secondary	-1.325	***	0.266	[0.214 0.330]
	(Ref: Higher education)				
Living Area	(Sub)urban	0.678	***	1.971	[1.691 - 2.297]
	(Ref: Rural)				
Nagelkerke R^2 = 0.1074					

sensitive to the presence of an interviewer are considered, namely those regarding work satisfaction, political interest and attitudes towards immigrants.

Factual questions Table 3 compares the answers from the online access panel respondents and the face-to-face interviewees to two 'factual' questions for which social desirability mode effects can be considered negligible. The propensity-weighted results for these variables are only shown to illustrate the effect of the propensity score adjustment. Since these covariates are used in estimating the probability of participation in the online access panel survey (cf. Table 2), their distributions are adjusted to be similar for both samples. This means that the online panel respondents are matched with those randomly selected in the general population with regard to their propensity to participate in the online access panel survey. As a consequence, the percentage distributions of living area and work participation do not differ significantly after applying this weighting technique.

From Table 3 it can be seen that significantly more online panel respondents live in an urban or suburban environment (this can be a big or small city or the suburbs of a big city), compared with the randomly-sampled persons who rather live rurally. Large differences in percentage points between both survey methods are observed (17%). After weighting for age, gender and education, these significant differences between the two samples do not disappear. The post-stratification weighting using these basic characteristics thus cancels out the under- and overrepresentations of the combined classes of age, gender and education, but does not make the online panel respondents and the face-to-face interviewees comparable regarding their living area. This illustrates that post-stratification can solve the problem of proportionality but not necessarily that of a (lack of) representativeness.

Since this factual question about people's living area can not be considered sensitive to social desirability, the observed differences are more likely to be related to actual respondent characteristics than to mode differences. The non-coverage issue is relevant here, since people living in cities have greater and easier access to the Internet compared with people living in more rural areas (Lenhart 2003:4). Thus, a possible explanation for this significant difference in responses is a selective non-response bias, namely that people from (sub)urban areas are more likely to be online and therefore more likely to be recruited to the online access panel.

Regarding work participation, Table 3 shows that substantially more respondents in the random sample have a paid job compared with respondents of the online access panel. This is rather counter-intuitive given the Internet penetration figures which show the opposite, namely that more Internet users have a paid job compared with people who do not have access to or do not use the Internet frequently (SCV survey 2006). Therefore, the expectation that online panels are dominated by actively working people is not found for the online access panel studied. Thus it can be noted that the online panel respondents (middle aged, without a paid job) are not necessarily the same as the group of frequent Internet users (young, having a paid job). The post-stratification weighting for age, gender and education does not make this difference between the two survey modes disappear. On the contrary, the difference in percentage points increases from 4 when unweighted to 6 when weighted.

Attitudinal questions about work satisfaction Next, only the respondents who claim to have a paid job are compared between both survey modes with regard to their work satisfaction. Table 4 illustrates that scaling the respondents' gen-

Table 3: Percentage distributions of living area and work situation

	Unweighted		Post-stratification weighted (age* gender* education)		Propensity weighted
	Online Panel	F2F Survey	Online Panel	F2F Survey	Online Panel
	%	%	%	%	%
Living area					
(Sub)urban	52.48	35.41	53.61	36.21	36.03
Rural	47.52	64.59	46.39	63.79	63.97
Chi^2 (df=1); p	87.46; ***		90.34;***		0.12; n.s.
Work					
Paid work	55.56	59.80	56.09	62.02	58.59
No paid work	44.44	40.20	43.91	37.98	41.41
Chi^2 (df=1); p	5.47;*		10.71; ***		0.45; n.s.

eral (dis)satisfaction with their current job produces no significant differences between the online panel and the face-to-face survey. On the other hand, a similar eleven-point scale regarding satisfaction with the time balance between paid work and other activities differs significantly between the online panel respondents and the personally-interviewed people; web respondents are on average slightly more satisfied. The post-stratification weighting has only a minor impact on the data. Similar inconsistencies are found with the second pair of questions that are concerned with the frequency with which people find their job interesting or a cause of stress. The seven-point scales show, in the former case a significant difference, and in the latter case a similar response distribution between both sample groups.

If the mode effect of social desirability is taken as a possible explanation of the observed differences, the expectation is that the personally-interviewed respondents will show a higher satisfaction compared with those who answered the questionnaire online. There is, however, no clear reason why this would play a part for some of the work-related questions (e.g. how often work is interesting) but not for others (e.g. how often work causes stress, general work satisfaction) or why occasionally the exact opposite of the expectations is found (e.g. time balance satisfaction). This suggests that although two different modes are compared, the mode differences certainly cannot always explain the differences in answers between the respondent groups.

The propensity-weighted means of the online panel survey only slightly change the survey results, and in the opposite direction to what would be expected. By making the distribution for several basic variables equal across both surveys based on propensity scores, it is expected that the differences between the propensity score adjusted panel data and the personally-interviewed respondents would become smaller. This seems not to be the case here, since the differences have become slightly larger instead of smaller after applying this weighting technique.

Attitudinal questions about political interest and complexity Another theme that lends itself to studying the possible selection bias is political interest and political complexity. In previous research, it has been found that online panel respondents show more politically-active attitudes (Duffy et al. 2005:627). Accordingly, it can be expected that they are more interested in politics and find politics less complicated in comparison with the randomly-sampled respondents. Indeed, Table 5 shows that more online panel respondents answer the extreme category 'very interested in politics' compared with the personally-interviewed respondents of whom more answer 'not at all interested'. In the middle categories however, attitudes move in the opposite direction. Moreover, when the mean values of this variable are compared instead of the percentage distributions, the significant differences between both sample groups disappear. Thus, the suggestion that panel respondents are more interested in politics compared with the randomly-sampled persons has to be put into perspective.

With regard to the frequency with which respondents find politics so complicated that they cannot really understand it, and the degree of difficulty they have in forming a political opinion, it can be seen that online panel respondents more often tend to choose the middle categories, such as 'occasionally' and 'neither difficult nor easy'. This could be an indication of 'satisficing' from online respondents (Heerwegh and Loosveldt forthcoming). From the two questions concerning political complexity, only the last one shows a result in line with general expectations, namely that online panel respondents find it less difficult to make up their mind about political issues.

Weighting by age, gender and education only has a minor impact on the survey results, and does not make the online panel respondents comparable with the randomly selected people regarding their political attitudes. On average, the propensity score weighting changes the percentages slightly more than the post-stratification, although not always in the expected direction. By exploring possible explanations, it was found that the relationship between political in-

Table 4: Mean scale values concerning work satisfaction (only for those respondents who have a paid job)

	Unweighted		Post-stratification weighted (age* gender* education)		Propensity weighted
	Online Panel	F2F Survey	Online Panel	F2F Survey	Online Panel
	Mean	Mean	Mean	Mean	Mean
Q How satisfied are you with (scale 0 very dissatisfied 10 very satisfied)					
current work	7.654	7.637	7.661	7.621	7.656
t; p	0.19; n.s.		0.46; n.s.		0.21; n.s.
time balance paid job - other activities	6.922	6.268	6.958	6.201	6.979
t; p	6.50; ***		7.56; ***		7.00; ***
Q How often do you find that your job (scale 0 never 6 all the time)					
is interesting	4.401	4.591	4.392	4.596	4.350
t; p	-3.20; ***		-3.45; ***		-3.92; ***
causes stress	3.324	3.359	3.283	3.393	3.259
t; p	-0.46; n.s.		-1.44; n.s.		-1.29; n.s.

terest and education is not the same in both surveys.

Contrary to what would be expected, it seems that substantially fewer online panel respondents with a higher education degree are very, or quite, interested in politics compared with the more educated people in the random sample. On the other hand, the less educated online panel respondents more frequently answer that they are very interested in politics. For political complexity, there is no different relationship with education between both samples (the more educated, the less complex the respondent evaluates politics). Because of different underlying relationships, such as those between political interest and education, the propensity score weighting that is partially based on education makes the differences between both survey modes slightly larger instead of smaller. This indicates that the gap between both groups of respondents is becoming even larger after weighting. Therefore, the specificity of the online panel sample is an important explanation for the differences observed between both survey groups.

Attitudinal questions about immigrants Subsequent to the comparison of factual questions about living area and work participation, and attitudinal questions regarding work satisfaction and political interest, a theme is considered that is particularly sensitive to social desirability; attitudes towards immigrants. It can be assumed that the presence of an interviewer during the face-to-face survey produces more positive responses towards immigrants compared with those from the online panel respondents who answer the self-administered survey online.

Table 6 shows that all the mean scale values concerning attitudes towards immigrants are consistently higher for the online access panel respondents. This confirms the more negative attitudes towards immigrants in the online panel sample, compared with those from the personally-interviewed respondents. Also, when the percentage distributions of the first three ordinal variables are compared, it seems that the online panel respondents more often answer that only 'a few' immigrants or 'none' should be allowed to live in Belgium, compared with the face-to-face respondents who tend to answer in the categories 'allow some' and 'many'. This is in line with expectations based on the presence or absence of an interviewer during the survey.

As the available data do not allow for the disentangling of the mode effects and sample selection biases, it cannot be entirely ruled out that sample differences also play a part. Only a limited impact is observed due to the weighting based on age, gender and education, as well as the propensity score adjustment. The latter even makes the differences between the online panel and random sample slightly larger. Therefore, both the social desirability and the selectivity of the online panel sample should be seen as possible explanations for the differences observed.

Discussion and conclusion

To study the representativeness of an online access panel sample and the impact of weighting procedures, the survey results collected with volunteer online access panel members were compared to a probability-selected sample of the general population, who were interviewed face-to-face. Regarding the time needed to collect the survey information, a large difference was observed between the two survey modes. Almost one third of the online panel members invited responded on the very first day. It would be very expensive and difficult, if not impossible, to obtain a similar rapidity in the fieldwork process of a face-to-face survey. Moreover, the

Table 5: Percentage distributions and mean scale values of political interest and political complexity

	Unweighted		Post-stratification weighted (age* gender* education)		Propensity weighted
	Online Panel	F2F Survey	Online Panel	F2F Survey	Online Panel
	%	%	%	%	%
Q How interested are you in politics?					
1. Very interested	11.20	8.16	10.90	8.60	11.60
2. Quite interested	27.95	41.12	28.55	42.31	30.09
3. Hardly interested	43.66	29.59	43.92	27.47	43.23
4. Not at all interested	17.19	21.12	16.64	21.62	15.08
Chi^2 (df=3); p	91.51; ***		108.94 ; ***		86.86 ; ***
Mean	2.669	2.637	2.663	2.621	2.618
t; p	0.97; n.s.		1.29; n.s.		-0.58; n.s.
Q How often does politics seem so complicated that you cannot really understand what is going on?					
1. Never	3.29	5.74	3.38	5.56	2.75
2. Seldom	21.91	22.54	21.89	23.48	20.15
3. Occasionally	39.88	34.43	40.09	33.81	40.17
4. Regularly	25.46	23.98	25.30	23.39	26.86
5. Frequently	9.47	13.32	9.33	13.77	10.07
Chi^2 (df=4); p	28.62; ***		32.50; ***		36.87; ***
Mean	3.159	3.166	3.153	3.163	3.214
t; p	-0.19; n.s.		-0.28; n.s.		1.29; n.s.
Q How difficult or easy do you find it to make your mind up about political issues?[a]					
1. Very easy	4.34	2.97	4.50	2.70	3.91
2. Easy	23.06	21.29	23.05	22.28	22.15
3. Neither difficult nor easy	43.04	33.67	42.65	32.81	43.56
4. Difficult	23.93	35.31	24.26	34.92	25.12
5. Very difficult	5.63	6.76	5.54	7.29	5.26
Chi^2 (df=4); p	58.08 ; ***		58.99 ; ***		50.19; ***
Mean	3.035	3.216	3.033	3.218	3.057
t; p	-5.27; ***		-5.36; ***		-4.68; ***

[a] The original answer categories are recoded for reasons of consistency, so that higher scores represent a higher political complexity, in accordance with the other questions in Table 5.

costs of conducting an online panel survey compared with those of a face-to-face interview are much lower. Such time and cost reductions make online panel surveys a very attractive data collection method.

However, the sample obtained by the online panel selection approach deviated significantly from the general population regarding age, even though this variable was used for pre-stratification. In particular, the oldest (65-74 years old) and youngest (18- 24 years old) age categories were underrepresented in the panel sample, which is to be expected for the first group but not for the second group of respondents, who are generally considered as Internet conversant. Another remarkable observation concerns the group of people between 55 and 65 years old, who were substantially overrepresented in the panel sample, which could indicate that this group not only has sufficient time but also the necessary skills for participating in online panel surveys. This illustrates the importance of permanently monitoring the groups of Internet users and their Internet usage to select relevant weighting variables. As changes in Internet usage can be expected in the future, the participation in online access panel surveys may change accordingly.

By comparing the data collected from online access panel members with results obtained through a face-to-face survey of randomly sampled members of the general population, significant differences were observed in almost all of the response distributions for different themes such as work satisfaction, political attitudes and ethnocentrism. In explaining the observed differences, both the specific characteristics of the respondents in the online access panel and the face-to-face survey as well as the features of the different modes have to be considered. For the factual questions about people's living area and work situation, significant differences were observed. Since mode effects probably do not affect these questions, it suggests selectivity in the online access panel sample. The individuals recruited by different channels

Table 6: Percentage distributions and mean scale values about attitudes towards immigrants

	Unweighted		Post-stratification weighted (age* gender* education)		Propensity weighted
	Online Panel	F2F Survey	Online Panel	F2F Survey	Online Panel
	%	%	%	%	%
Q To what extent do you think Belgium should allow people of *the same race or ethnic group* as most Belgian people to come and live here?					
1. Allow many	13.13	18.51	13.57	18.06	12.4
2. Allow some	52.26	56.85	51.99	57.92	52.06
3. Allow a few	26.80	18.40	27.02	17.40	27.07
4. Allow none	7.81	6.24	7.42	6.62	8.46
Chi^2 (df=3); p	41.05; ***		43.04; ***		49.22; ***
Mean	2.293	2.124	2.283	2.126	2.316
t; p	5.83; ***		5.41; ***		6.58; ***
Q How about people of a *different race or ethnic group* from most Belgian people?					
1. Allow many	8.53	10.13	9.32	9.65	7.58
2. Allow some	43.24	49.03	41.99	51.07	41.62
3. Allow a few	33.84	26.41	35.15	24.92	35.18
4. Allow none	14.40	14.43	13.54	14.36	15.62
Chi^2 (df=3); p	20.34; ***		37.55; ***		32.73; ***
Mean	2.541	2.451	2.529	2.440	2.588
t; p	2.88; **		2.86; **		4.39; ***
Q How about people from *the poorer countries* outside Europe?					
1. Allow many	8.19	10.33	9.29	10.20	7.77
2. Allow some	44.08	44.48	42.64	45.88	42.53
3. Allow a few	30.78	31.60	31.36	31.01	31.94
4. Allow none	16.95	13.60	16.72	12.91	17.76
Chi^2 (df=3); p	9.16; *		9.19; *		13.87; **
Mean	2.565	2.485	2.555	2.466	2.597
t; p	2.52; **		2.77; **		3.51; ***
	Mean	Mean	Mean	Mean	Mean
Immigration improves or worsens (Scale 0 improving 10 worsening)[b]					
economy	6.544	6.275	6.500	6.230	6.675
t; p	2.87; **		2.87; **		4.31; ***
cultural life	5.908	5.249	5.836	5.243	6.015
t; p	6.53; ***		5.86; ***		7.67; ***
place to live	6.843	6.193	6.771	6.186	6.938
t; p	6.75; ***		5.72; ***		7.77; ***

[b] These three scales are, for reasons of consistency, recoded so that higher values indicate a more negative attitude towards immigrants.

to join the online panel and to regularly respond to surveys seem to be a specific group of people who deviate in these characteristics from the general population. The observation of Duffy et al. (2005) that online panel respondents are more politically active could not be found in this particular study. Although it was observed that online panel respondents find it less difficult to form an opinion about politics compared to the general population, the suggestion that online panel respondents are more often interested in politics has to be put into perspective.

Additionally, a relevant mode effect has to be mentioned when the answers from an online survey are compared with a face-to-face survey. The presence of an interviewer during the face-to-face survey could make people respond differently than the respondents who answer the questionnaire through the Internet. For the questions regarding work sat-

isfaction, rather diverse results were found, which were not always in line with the social desirability effect. For a theme that can be considered more sensitive to this mode effect, attitudes towards immigrants, the survey results are consistently in agreement with the different social distance between interviewer and respondent. The more negative attitudes towards immigrants living in Belgium can at least partly be explained by the difference in social desirability in both survey modes. But this effect is considered to play a role in addition to the sample selectivity. Due to data limitations, however, it is not possible to give a more definite conclusion about the relative impact of both.

Post-stratification weights based on age, gender and education did not have a substantial impact on the results. This kind of weighting technique makes the proportions of the variables used comparable, but this does not necessarily make the answers of web respondents and personally-interviewed people more comparable with regard to attitude questions. A possible solution is the propensity score adjustment of the online panel data (Duffy et al. 2005; Taylor 2005). However, in this particular study the application of this weighting technique resulted in only minimal changes, and the differences between the respondent groups sometimes became larger instead of smaller. Although it can be argued that other and more variables need to be used for the post-stratification weighting and propensity score adjustment, this study illustrates that weighting methods using basic variables do not make the online panel data more comparable with the general population. This implies that non-probability based online panels should be used with caution. Only when representativeness is not required, such as for exploratory studies or experimental tests, online access panels might be used with more confidence (Bethlehem and Stoop 2007:127; Malhotra and Krosnick 2007:312).

As the present study is limited in attributing the observed differences to differential mode effect or sampling selectivity, future investigation is required to disentangle the mode and sampling effects. A standard procedure should be established for evaluating the representativeness of the online access panel sample, for example by comparing the online panel data with 'real world' figures for which census data are available (e.g. election results, urbanization and residence types). Additionally, further research is needed to identify the impact of the online panel survey approach on the relationships between variables.

Acknowledgements

An earlier version of this article was presented at the second European Survey Research Association (ESRA) Conference, 25-29 June 2007 in Prague, Czech Republic. We are grateful for comments from participants at this meeting and from the anonymous referees. This research was supported by a PhD fellowship from the Research Foundation Flanders (FWO) for the second author (Nathalie Sonck).

References

Bandilla, W., Bosnjak, M., & Altdorfer, P. (2003). Survey Administration Effects? A Comparison of Web-Based and Traditional Written Self-Administered Surveys Using the ISSP Environmental Module. *Social Science Computer Review*, *21*, 235-243.

Bethlehem, J., & Stoop, I. (2007). Online panels - a Paradigm Theft? In M. Trotman et al. (Eds.), *The challenges of a changing world* (p. 113-131). Southampton, UK: Association for Survey Computing.

Comley, P. (2003). *Innovation in Online Research - Who Needs Online Panels?* Research2003. (Conference Paper 36, 14p.)

Couper, M. P. (2000). Web Surveys. A Review of Issues and Approaches. *Public Opinion Quarterly*, *64*(4), 464-494.

de Leeuw, E. D. (1992). *Data Quality in Mail, Telephone and Face-to-Face Surveys*. Amsterdam: TT-Publicaties.

de Leeuw, E. D. (2005). To Mix or Not to Mix; Data Collection Modes in Surveys. *Journal of Official Statistics*, *21*, 233-255.

Dillman, D. A. (2000). *Mail and Internet Surveys. The Tailored Design Method*. New York: John Wiley & Sons.

Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing Data from Online and Face-to-Face Surveys. *International Journal of Market Research*, *47*(6), 615-639.

Heerwegh, D., & Loosveldt, L. (n.d.). *Face-to-Face versus Web Surveying in a High Internet Coverage Population: Differences in Response Quality.* (Public Opinion Quarterly, forthcoming)

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires; Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly*, *67*, 79-125.

Kellner, P. (2004). Can Online Polls Produce Accurate Findings? *International Journal of Market Research*, *46*(1), 3-21.

Lee, S. (2006). Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys. *Journal of Official Statistics*, *22*(2), 329-349.

Lenhart, A., Horrigan, J., Rainie, L., Allen, K., Boyce, A., Madden, M., et al. (2003). *The Ever-shifting Internet Population; A New Look at Internet Access and the Digital Divide*. Washington DC: Pew Internet and American Life Project.

Malhotra, N., & Krosnick, J. A. (2007). The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Non-probability Samples. *Political Analysis*, *15*, 286-323.

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*, 41-55.

Schonlau, M., van Soest, A., & Kapteyn, A. (2007). Are 'Webographic' or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring? *Survey Research Methods*, *1*(3), 155-163.

Schonlau, M., Van Soest, A., Kapteyn, A., & Couper, M. P. (2006). *Selection Bias in Web Surveys and the Use of Propensity Scores.* Working Paper of RAND Labor and Population.

Schwarz, N., Strack, F., Hippler, H.-J., & Bishop, G. (1991). The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology*, *5*, 193-212.

SCV survey. (2006). *Sociaal-Culturele Verschuivingen in Vlaanderen.* Available from `http://aps.vlaanderen.be/survey/aps_survey/APS-survey.htm`

Stenbjerre, M., & Laugesen, J. N. (2005, April). *Conducting Representative Online Research; A Summary of Five Years of Learn-*

ing (zapera). (paper presented at ESOMAR Worldwide Panel Research Conference in Budapest)

Taylor, H. (2005). Does Internet Research 'Work'? Comparing Online Survey Results with Telephone Surveys. *International Journal of Market Research*, *42*(1), 51-63.

Van Ossenbruggen, P. W., & Vonk, T. (2006, October). *Nederlands Online Panel Vergelijkingsonderzoek (nopvo).* Available from `http://www.dans.knaw.nl` (Presentation of DANS Symposium 'Access Panels en Online Onderzoek, Panacee of Slangenkuil' in Amsterdam)

Vehovar, V. (2002). Nonresponse in Web Surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (p. 229-242). New York: J. Wiley & Sons.

Vehovar, V., Batagelj, Z., & Lozar, K. (1999, May). *Web surveys: Can the Weighting Solve the Problem?* (Paper presented at the AAPOR Conference in St. Petersburg)