Discussion Paper

Establishing the accuracy of online panels for survey research

2016 | 04

E. Brüggen (Maastricht University School of Business and Economics)
J. van den Brakel (Statistics Netherlands)
J. Krosnick (Stanford University)

Summary

Many surveys being conducted today for academic research, government policy-making, and marketing collect data via the Internet from groups of respondents who volunteered to answer questions regularly, rather than from random samples of individuals who were selected using the scientific methods that have dominated survey research for decades. This paper compares the accuracy of results obtained from 18 such opt-in online "panels" with the results obtained from respondents selected randomly from the population who answered questions either via the Internet or via face-to-face interviewing. The non-probability samples yielded less accurate estimates of proportions and notably different relations between variables than did the probability samples, and these differences were not eliminated by weighting. These findings reinforce the value of scientific, random sampling to permit generalizing research findings to a larger population. These findings suggest that the marketing community should pay more attention to and provide elaborate and honest descriptions of the nature of survey samples, to allow consumers of the data to assess their likely accuracy.

Index

page

1. Introduction

Online panels are increasingly being used in research on public opinion, marketing, psychological and social processes, and medical phenomena (Baker et al. 2010). ESOMAR reported that in the 2014 global market research report that $11.3billion were spent on online research, the vast majority of which involved opt-in online panels assembled via methods that do not involve random sampling from the population of interest. Such opt-in panels consist of a large number (often purportedly tens of thousands or even millions) of people who responded to advertisements or other efforts to invite them to volunteer to complete surveys regularly, with no known probability of selection from the population. Popular methods for online panel respondent recruitment include partnerships with commercial Internet vendors (co-registration agreements), e-mail marketing/mass emailing (email list brokers), affiliate hubs, display ads or banner advertising, snowballing, text links, and pop-up, website intercepts (Postoaca 2006). The popularity of this new approach to generating survey samples is tremendous and has often been attributed to claims of faster turnaround time, easier access of data to researchers, and lower costs than is permitted by probability sampling (Baker et al. 2010).

Despite the practical advantages of opt-in online panels, their accuracy should be a key concern for survey researchers. When researchers seek to generalize their findings to a population, inexpensive and quick survey results may not be worth their apparent benefit if they are notably less accurate than results obtained via scientific, probability sampling. In fact, buyers of data from opt-in samples have voiced concerns about accuracy. In 2006, Kim Dedeker, then P&G vice president of global consumer and market knowledge and current chair of Americas for Kantar (a division of WPP), caused an uproar in the research community with her critical comments about the lack of quality of findings from opt-in online panels, which she mainly attributed to poor representation of the population of interest in these panels (http://www.rflonline.com/clientsummit/notes/Kim-Dedeker-notes.html).

In theory, opt-in samples may sometimes yield results that are as accurate as probability samples, assuming that the factors that explain a population member's presence or absence in the sample are uncorrelated with variables measured in a study and the magnitudes of associations between pairs of variables. However, a growing number of studies have shown that opt-in panels routinely over- or under-represented a series of population subgroups (Chang and Krosnick 2009; Couper 2000; Dever et al. 2008; Malhotra and Krosnick 2007) and that opt-in samples yielded less accurate results than did probability samples (Chang and Krosnick 2009; Yeager et al. 2011). However, these studies have been conducted almost exclusively in the U.S. and have often examined a relatively small set of opt-in panels. So additional studies, especially ones examining sizable numbers of panels outside the U.S., would be of value.

Firms providing data from opt-in panels sometimes argue that sophisticated weighting techniques can solve problems of sample unrepresentativeness (e.g., Harris Interactive (http://www.harrisinteractive.com/partner/methodology.asp) proposed "propensity score weighting" for this purpose). Although weighting techniques are well-developed and widely used with probability samples (e.g., Frölich 2004; Heckman et al. 1998; Ichimura and Taber 2001; Särndal et al. 1992), claims about the benefits of using such techniques with opt-in samples have not been sustained by the few studies done to date on this issue. Some studies found that propensity score weighting decreased but did not eliminate selection biases that afflict opt-in panel estimates, and the reduction in bias came at the cost of considerably increased variance (e.g., Lee 2006; Valliant and Dever 2011; Yeager et al. 2011). Other studies found weighting more often reduced accuracy rather than improving it (Yeager et al. 2011). Nonetheless, this issue also merits further investigation.

Remarkably, despite the great interest across the social and behavioral sciences in studying relations between variables, almost no evidence to date has explored the accuracy of relations between variables as gauged using data from opt-in samples. Even if distributions of variables gauged by opt-in samples are not as accurate as those yielded by probability samples, the relations between variables may nonetheless be very similar (e.g., Kivlin 1965). That is, if a researcher aims to assess the relation of gender to life satisfaction, it is not in and of itself problematic if the sample over-represents females. If, however, low life satisfaction is associated with increased opt-in survey participation among females but with decreased survey participation among males, then researchers would reach an erroneous conclusion about the association between gender and life satisfaction. This issue also merits careful study.

We set out to contribute to the literature by generating new evidence about (1) the accuracy of distributions of variables gauged via opt-in samples and probability samples interviewed by the Internet and face-to-face using data collected outside the U.S. from a large number of opt-in panels; (2) the accuracy of *relations* between variables observed with data from opt-in samples; and (3) whether *weighting* improves the accuracy of results generated with opt-in samples.

To this end, we analyzed data from 18 non-probability online panel firms that participated in the Dutch Online Panel Comparison Study (called NOPVO; http://www.nopvo.nl), one probability sample of residents of the Netherlands who provided data via the Internet (the LISS panel; Scherpenzeel 2009), and two probability samples of Dutch residents interviewed via computer assisted personal interviewing (CAPI) by Statistics Netherlands. To gauge accuracy, we compared measurements from those data sources with benchmarks from the Dutch government's registry of all 16 million residents of the country: the Municipal Basic Administration (MBA).

Thus, in contrast to previous studies that have almost exclusively assessed accuracy using benchmarks

constructed from high-quality probability sample government surveys, the benchmarks used here entail no sampling error or selection bias: Dutch residents are required by law to contribute data to municipal population registers, which the Dutch Ministry of the Interior and Kingdom Relations has said are 97.8% accurate (http://www.bprbzk.nl/GBA/Kwaliteit). The Netherlands is also a particularly good test bed for this investigation, because the extremely high rate of Internet adoption in the country (94%) means that firms soliciting opt-in samples confront minimal non-coverage barriers (Seybert 2012).

Because more opt-in panel firms provided data for this study than for any prior study released publicly, it was possible to explore whether the accuracy of results obtained by a survey was related to techniques used to manage the panels and sometimes claimed to promote accuracy. Some of the considerations we examined are panel recruitment strategies: recruitment via links on websites, from research conducted through traditional data collection methods, emails sent to addresses purchased from other companies, recommendations of potential panel members by existing panel members, telephone recruitment, or invitations sent to members of existing mail panels. We also explored whether accuracy was related to panel management attributes, such as the number of active members of a panel (a larger panel is often asserted to yield more accuracy), the number of survey invitations sent to panel members per month (fewer invitations are often asserted to yield more accuracy), the rate of panel member drop out per year (a higher dropout rate is often asserted to yield a fresher, less fatigued panel), and whether the panel is supplemented with newly recruited individuals (which is thought to enhance panel representativeness). Finally, we explored whether accuracy is related to the type of rewards offered to panel members in exchange for their registration (use of reward for registering in the panel) and participation (cash, points redeemable for cash or prizes, lotteries, or lottery tickets). We also examined whether non-probability sample Internet survey firms that had been in business for more years yielded higher accuracy.

In doing all this, we confronted an important statistical challenge. In past studies, the numbers of respondents in the various survey samples were essentially equal (e.g., Yeager et al. 2011), so expected accuracy was equivalent as well. However, the probability samples examined here are substantially larger than the sizes of the opt-in Internet survey samples. Therefore, based on sample size alone, one might expect the probability samples to be more accurate.[1] Furthermore, conventional statistics, such as a Student's t-test or a Welch t-test, for testing whether the probability samples are significantly more accurate than the non-probability samples will be inclined toward rejecting the null hypothesis because of the huge sample sizes of some probability samples and cannot distinguish whether the observed differences in accuracy are the result of the differences in sample size or of the sample selection mechanism. Therefore, we propose a new analysis method that levels the differences in sample sizes across samples.

Survey data contain errors that stem from various sources. Generally distinction is made between

sampling errors, which arise since only a part of the target population is observed, and non-sampling errors. Under a good sampling strategy, i.e. the right combination of sample design and estimator, the estimator is (approximately) design-unbiased and sampling errors do not have a systematic effect on the estimates. Non-sampling errors, on the other hand generally result in biased estimates for the population parameters. One possible classification is to distinguish between measurement bias and selection bias. Measurement bias implies that the answers obtained from the respondents are obscured by errors, which might have a systematic effect. The amount of measurement bias typically depend on questionnaire design, wording of the question and the interviewer mode. Selection bias arise since a part of the population does not respond or cannot be reached in the sample due to under coverage in the sample frame or the use of an interviewer-mode that does not cover the entire target population.

As we will explain in more detail in the next paragraph, our research relies on different data sources which allow us to compare the accuracy for different modes (online vs. face-to-face) as well as different sampling techniques (probability versus non-probability). The purpose of this study is to analyze the amount of selection bias in voluntary opt-in sample, which arise since a non-probability mechanism is used to select the samples. To avoid confounding with measurement error, we carefully selected variables that are less sensitive for measurement error and mode effects and which were measured identically or nearly identically[3] in all data sources.

2. Data

To answer our research question, we rely on data obtained from the MBA, two probability samples of the Statistics Netherlands (Labour Force Survey and Permanent Survey on Living Conditions), a Web panel that is based on a probability sample and 18 non-probability opt-in panels. These data sources allow us to compare online and offline as well as probability and non-probability samples.

2.1 Non-Probability Internet Panels (NOPVO)

All online market research panels in the Netherlands were invited to participate in this study, called NOPVO, and 95% of those companies (19) did so. One of the firms failed to provide measurements of many of the required variables, so the comparisons reported here focus on the remaining 18 non-probability panels. All participating panels were asked to collect data from 1,000 respondents who were representative of the Dutch population ages 18 to 65. Across companies, all panelists were invited to complete the same survey on the same day, and the questionnaire was scripted and hosted centrally by an independent party. Data collection took place over the seven-day period from April 20 to 27, 2006.

The panel management techniques used by the firms are described in Table 1 in the appendix. Across panels, the average response 500, but it differed widely across panels and ranged from 184 to 769 respondents (see Table 3 in the appendix).

2.2 Probability Internet Panel

Supported by a grant from the Dutch National Science Foundation, CentERdata at Tilburg University created the LISS panel (Longitudinal Internet Studies for the Social Sciences) by drawing a simple random sample of addresses from the Dutch population register. Then, all household members residing at a sampled address were invited by letter, telephone call, or household visit to join the panel. People without a computer or Internet connection were given the necessary equipment and access to the Internet with a broadband connection. The final panel participation rate is 48% of the total invited sample, or 9,844 households (for more information, see Scherpenzeel 2009). We used data from their so-called “background survey” conducted in February, 2008, once the recruitment process was finished and the panel was at full strength.

2.3 Face-To-Face Probability Sample Surveys

One probability sample of Dutch residents that we examined was interviewed face-to-face by Statistics Netherlands (the national statistical agency): the so-called Labor Force Survey (LFS). Stratified two-stage

random sampling of addresses was done from the population registry (called the MBA). All households residing at sampled addresses were included in the sample. All household members aged 15 and older were included in the sample. If a household member could not be interviewed directly, another household member was permitted to provide proxy reports on the individual. In 2006, 10,589 households provided data, and in 2008, 10,632 households provided data - about 65% of the approached households.[2]

We also analyzed data from the Permanent Survey on Living Conditions (PSLC), another face-to-face Statistics Netherlands study based on a stratified two-stage random sample of persons residing in the Netherlands, ages 15 years and older. About 60% of the persons in the sample completed interviews. In 2006, 9,607 persons provided data, and in 2008, 9,499 persons provided data.

2.4 Municipal Basic Administration (MBA)

Dutch citizens are required by law to report changes in their demographics to their municipalities. All changes must be reported in person, by Internet, or by mail within 5 working days, and residents must present a valid ID (e.g., passport, driver's license, or Dutch identity card) and proof of the change (e.g., a rental contract or proof of house ownership, an official birth certificate from a doctor or midwife). The MBA constitutes a very strong source for benchmarking, since it is highly accurate (http://www.bprbzk.nl/GBA/Kwaliteit) and entails no notable sampling error or selection bias. See Bakker (2012) for a more detailed assessment of the quality of the MBA.

3. Analysis

We compared the surveys to the MBA using all variables that were measured identically or nearly identically[3]: gender, age, urbanization, country of origin of the respondent, country of origin of the respondent's parents, province of residence, region of residence, and the number of people living in the household (which we call the "register variables"). We also compared the Internet samples to the probability face-to-face samples in terms of two additional variables not in the MBA: education and employment (which we call the "non-register variables"). The Benchmark for employment was measured in the LFS. The benchmark for education was measured in the PSLC. Finally, two other non-register variables, health quality and life satisfaction, were measured identically in the PSLC and in the non-probability Internet surveys, thus permitting additional comparisons, using the PSLC as the benchmark. The exact wording of all questions and response categories can be found in Appendix A.

3.1 General Regression and Horvitz-Thompson Estimators

Point and variance estimates for all surveys were obtained using the Horvitz-Thompson (HT) estimator for comparisons of variables in the register and the general regression (GREG) estimator for variables not in the register. Let *N* denote the size of the target population and *n* the sample size. Both estimators can be expressed as the sum over the weighted observations obtained in the sample, i.e.

$$\hat{t}_y = \sum_{i=1}^{n} w_i y_i, \quad (1)$$

with y_i, the observation obtained from sampling unit i, and w_i, a weight that is determined so that (1) is an approximately design-unbiased estimate for the unknown population total $t_y = \sum_{i=1}^{N} y_i$. The HT estimators (Horvitz and Thompson 1952) use the so-called design weights, which are defined as the inverse of the probability that a sampling unit is included in the sample, to account for unequal selection probabilities in the LFS and PSLC. For example, in the case of simple random sampling, each unit has the same inclusion probability n/N. As a result, each element has the same design weight $w_i = N/n$, and (1) reduces to the unweighted sample mean multiplied with the population size.

The GREG estimator (Särndal et al. 1992) attempts to improve the accuracy of the HT estimator by taking advantage of auxiliary variables for which the population totals are known *a priori* from the MBA. To this end, the GREG estimator calibrates the design weights such that the sum over the weighted auxiliary variables in the sample is exactly equal to their known population totals. More formally,

$$\hat{\boldsymbol{t}}_{\boldsymbol{x}} = \sum_{i=1}^{n} \widetilde{w}_i \boldsymbol{x}_i = \boldsymbol{t}_{\boldsymbol{x}}, \quad (2)$$

where $\boldsymbol{x}_i$, is a *Q* vector with auxiliary variables of unit *i* and $\boldsymbol{t}_{\boldsymbol{x}}$ denotes a corresponding vector with the

population totals of the auxiliary variables that are included in the weighting scheme of the GREG estimator. Finally, $\widetilde{w}_i$ are the calibrated weights obtained with the GREG estimator, see Särndal et al. (1992), Ch. 6 for a general expression. The GREG estimator is motivated with a linear regression model, which defines the relation between the target variable and the auxiliary variables, i.e.

$$y_i = \boldsymbol{\beta}^t \boldsymbol{x}_i + e_i, \qquad (3)$$

with $\boldsymbol{\beta}$ a *Q* vector with regression coefficients and e_i a residual of unit *i.* If this linear regression model explains the variation of the target variable reasonably well, then the GREG estimator reduces the variance of the HT estimator and corrects, at least partially, for selective nonresponse.

A well-known special case of the GREG estimator is poststratification. In this case $\boldsymbol{x}_i$ is a categorical variable which divides the population in *Q* subpopulations of size N_q, *q=1, …, Q*. In this case it can be shown that (1) equals $\hat{t}_y = \sum_{q=1}^{Q} \sum_{i=1}^{n_q} \frac{N_q}{n_q} y_i \equiv \sum_{q=1}^{Q} N_q \bar{y}_q$, which can be recognized as the well-known poststratification estimator, Särndal et al. (1992), Ch. 7.6. If more than one categorical auxiliary variable is available, the poststratification estimator is obtained by the complete cross-classification of these auxiliary variables and requires the availability of the joint population frequencies. The obvious drawback is that the number of sampling units in the poststrata can become too small or even zero, which makes the poststratification estimator unstable. One solution is to collapse cells with small or zero sample sizes. Another, more general, possibility is to include the auxiliary variables in the linear model (3) underlying the GREG estimator and skip higher order interaction terms and weight to the marginal population frequencies.

Large deviations between the distribution of the auxiliary variables in the sample and the population can result in negative GREG weights. This problem is sometimes circumvented by applying iterative proportional fitting. This is a related approach where weights are calibrated such that the weighted auxiliary variables in the sample meet requirement (2) while negative weights cannot occur by definition. Less appropriate weighting models that result in negative weights under the GREG estimator, generally result in undesirable increase of the weights under iterative proportional fitting. A drawback of iterative proportional fitting is that no analytic expressions for the variance of the sample estimates is available. Therefore the GREG estimator is applied in this paper.

It follows from (2) that the GREG estimator is not useful for the analysis of the register variables, since they are used as auxiliary variables in the weighting procedure and therefore, by definition, exactly equal the known population value. Therefore, the GREG estimator is only used for the non-register variables, and the HT estimator is used for the analysis of the register variables.

The NOPVO panels are not based on a probability sample, which hampers the derivation of design weights. For these panels the HT-estimator stands for calculating the unweighted sample means, which

comes down by assuming simple random sampling. With the GREG estimator, the design weights are also assumed to be equal for all elements in the NOPVO panels, and auxiliary information is used in attempt to improve the accuracy of the estimates.

For each sample (NOPVO, LISS, LFS, and PSLC), an optimal weighting model was constructed for the non-register variables. The weighting models for the LFS and PSLC are based on the models used by Statistics Netherlands for the analysis of these surveys. For LISS and the non-probability Internet panels, weighting models were constructed by predicting the target variables in a linear regression and using a step-forward procedure to select predictors from among all available auxiliary variables. The set of potential auxiliary variables contained Age(5), Gender(2), Household size(6), Nationality(3), Province(12), and Urbanization degree(5). All variables are categorical and the number between brackets denote the number of categories. In the variable selection procedure also all two-way interaction terms between these variables are considered. This resulted in weighting models that explain as much of the variation in the target variable as possible. Then, the GREG estimator was applied to construct an initial set of weights. To find the optimal weighting model for each panel, we removed interaction terms and auxiliary variables until we excluded negative weights and over-dispersion of the weights and reached a minimum number of 10 responses in each cell.

The weighting models used for each source are listed in Table 2 in the appendix. Particularly for the non-probability internet panels sparse weighting models are selected to avoid negative weights and problems with empty cells. This is not only the result of the small sample sizes but is also caused by the fact that these samples are very skewed with respect to the distribution of the auxiliary variables in the target population. For five non-probability internet panels, negative weights could not be avoided, which are obviously problematic and reveal the substantial challenge inherent in building effective weights for such datasets. In our final analyses, the negative weights were circumvented by applying the bounding algorithm of Huang and Fuller (Huang and Fuller 1978).

3.2 Accuracy Assessment

The target variables considered in this paper for the accuracy evaluation are all categorical. In keeping with past research, we gauged the accuracy of the various data sources by comparing the estimated proportions of the modal response category in the samples to each question to the proportions of people in that category in the benchmark (Yeager et al. 2011). The benchmark for register variables is the MBA. For the non-register variables, the GREG estimates obtained with the LFS or PSLC are used as a basis for comparison. Then, we gauged whether weighting improved the accuracy of the survey samples. We report the absolute difference between each survey's result and its benchmark, as well as the

average of the absolute modal differences across questions.

Two problems arise when comparing the probability samples with the non-probability Internet panels. The first problem is that there is no known probability mechanism to select the sampling units of the non-probability samples. As a result, strong assumptions about the distribution of the sample statistics must be made to make accuracy statements about the sample estimates or to test whether the sample statistics are significantly different from the known benchmark. Put more simply: there is no scientific basis for calculating proper standard errors and confidence intervals for statistics computed with the non-probability sample data, see also Baker et al. (2010). In contrast, it is routine to compute proper standard errors for the probability samples, because they were produced using explicit randomization mechanisms to select the sample units from a finite target population, for which the distributions of the sample statistics are known. The second problem is the large differences in sample sizes between the non-probability Internet samples and the probability samples. The latter are much larger and therefore have considerably smaller sampling errors. Therefore, smaller deviations from the benchmarks can be expected for the large probability samples.

To cope with both problems, the following analysis procedure is developed (based on a suggestion from Professor B. Efron, Stanford University). All variables considered in this paper have multinomial responses which are transformed to proportions of units classified in $K \geq 2$ categories. The analysis is restricted to differences in the modal category. Let P_B denote the proportion of the modal category of the benchmark. For register variables, this value is obtained from the MBA. For non-register variables, it is the GREG estimate observed with the probability samples of Statistics Netherlands. Let $\hat{P}_{pr}$ denote the sample estimate for this proportion based on the probability sample of size n_{pr}. Furthermore $\hat{P}_{nopvo,x}$ denote the sample estimate for the same proportion based on the *x*-th NOPVO sample of size $n_{nopvo,x}$. The absolute deviations of the two sample estimates from the benchmark value are defined as $\hat{\Delta}_{pr} = \left|\hat{P}_{pr} - P_B\right|$ and as $\hat{\Delta}_{nopvo,x} = \left|\hat{P}_{nopvo,x} - P_B\right|$ for the probability sample and the *x*-th NOPVO sample respectively. Our purpose is to test the hypothesis $H_0: \hat{\Delta}_{pr} = \hat{\Delta}_{nopvo,x}$ against the one-sided alternative $H_1: \hat{\Delta}_{pr} < \hat{\Delta}_{nopvo,x}$. This hypothesis is tested by assuming a binomial distribution for $\hat{\Delta}_{pr}$, which is reasonable, because it is the proportion of sampling units in the modal category that are erroneously classified differently. We calculated the probability that the observed difference $\hat{\Delta}_{nopvo,x}$ for each NOPVO sample is a realization from the binomial distribution of $\hat{\Delta}_{pr}$, using the sample size of the NOPVO sample. More precisely:

$$P\left(k \geq K_{nopvo,x}\right) = \sum_{k=K_{nopvo,x}}^{n_{nopvo,x}} \binom{n_{nopvo,x}}{k} \left(\hat{\Delta}_{pr}\right)^k \left(1 - \hat{\Delta}_{pr}\right)^{\left(n_{nopvo,x}-k\right)}, \qquad (4)$$

where $K_{nopvo,x} = n_{nopvo,x}\ \hat{\Delta}_{nopvo,x}$ denotes the observed difference between the number of sampling units in the modal category of the *x*-th NOPVO sample compared to the benchmark. By using the size of the NOPVO sample in this binomial distribution, the dispersion around $\hat{\Delta}_{pr}$ is based on the size of the

NOPVO sample. As a result, the test equates sample sizes across the non-probability Internet samples and the probability samples. Moreover, no assumptions about the distribution of $\hat{P}_{nopvo,x}$ are required, since we tested whether this observation can be considered to be a realization from the distribution in the probability sample. By doing so, approximations of the standard error of $\hat{P}_{nopvo,x}$ are circumvented. Equation (4) specifies the probability that under a binomial distribution with parameters $\hat{\Delta}_{pr}$ and $n_{nopvo,x}$ the difference between the number of sampling units in the modal category of the *x*-th NOPVO sample and the benchmark is at least as large as the observed difference $K_{nopvo,x}$. It can therefore be interpreted as the *p*-value for the test of the aforementioned hypothesis.

4. Results

4.1 Accuracy of Estimates

4.1.1 Register variables NOPVO.

The HT estimates for the register variables in the face-to-face probability sample surveys had average absolute differences to the benchmark of 2.57 and 2.00 (see the bottom row of Table 3 and Figure 1 in the appendix). The online probability sample survey (LISS) manifested an average absolute difference of 4.07, and the average absolute differences of the online non-probability samples range from 3.74 to 9.75.

Zooming into the details of table 3, we see that the panels with the smallest average absolute difference to the benchmark are panel 16 (3.74) and panel 15 (4.34). Panel 3 shows the largest average absolute difference to the benchmark (9.75).

When looking at the individual variables in table 3, remember that for the significance testing, we test if the difference in estimates of the NOPVO panels compared to the benchmark fit into the distribution of the differences of PSLC, LFS, or LISS with the benchmark. In terms of individual variables, the largest difference to the benchmark occurred for NOPVO panel 3, which substantially over-represented women (Δ benchmark: -21.65). Large differences between panels are apparent for the origin of the respondent, the origin of his/her mother, and the origin of his/her father. For origin of the respondent, all NOPVO panels show significant different results ($p<0.05$) compared to the differences with benchmark of LFS and PSLC. For origin of mother and origin of father, 34 and 33 of the 36 significance tests of NOPVO compared to LFS and PSLC show significant differences at $p<0.05$.

Across all NOPVO panels, province shows the smallest average absolute differences to the benchmarks (2.99), and 15 of the 36 comparisons of NOPVO compared to the differences of LFS and PSLC with the benchmark aresignificantly different at $p<0.05$. The proportion of significant differences compared to LFS and PSLC is very high for panel 2 (88% for both LFS and PSLC) and panel 3 (88% for LFS and 75% for PSLC), compared to panels with low proportion of significant differences such as panel 10 and 15 (both 50% for LFS and 63% for PSLC).

4.1.2 Resigister variables LISS

As mentioned before, the online probability sample survey (LISS) manifested an average absolute difference of 4.07. This average absolute difference is larger than for the two probability samples, but

smaller than the results for 17 of the 18 online non-probability panels. However, the significance test reveals that LISS differs significantly to the benchmark compared to LFS and PSLC. And this average absolute difference is only significantly lower than for 10 of the 18 NOPVO panels (see third but last row in table 3). Across all variables, the proportion of significant differences to the benchmark compared to LFS and PSLC is also rather high (88% for LFS and 75% for PSLC). The results for LISS differ significantly ($p<0.05$) to the benchmark compared to LFS and PSLC except for region (for both LFS and PSLC) and gender (only PSLC).

In sumary, the results reveal that panel 16 and 15 are the more accurate online non-probability panels, whereas panel 3 and 2 perform rather poorly since they show large average absolute differences to the benchmarks and a high proportion of significant differences for the individual variables. The offline probability samples were more accurate than the non-probability samples. LISS showed somewhat mixed results. For 81% of the variables the differences of LISS compared to the benchmark where significantly larger compared to the two probability samples. For NOPVO 73% of the variables the differences with the benchmark were significantly larger compared to the two probability samples. This is an indication that the large differences of the NOPVO variables are partially explained with the small samples and that the accuracy of LISS is somewhere in between the probability samples and the NOPVO panels.

4.1.3 Non-register variables – HT estimator NOPVO

For the non-resgister variables the GREG estimates of the probability samples are used as the benchmarks, since they are considered as the most reliable approximation of these population parameters. The results are summarized in Table 4 and Figure 1 in the appendix .

Comparing HT estimates for education and employment only, the non-probability Internet panels' average absolute differences compared to the GREG estimates of LFS or PSLC range from .73 to 15.58. Panel 8 shows the smallest average difference (0.73), which is even slightly smaller than LISS, followed by panel 9 (2.12). Panel 18 (15.58) and panel 14 (11.74) show the largest average absolute differences.

For all non-register variables, so including health and life satisfaction, the average absolute differences for NOPVO range from 1.88 to 9.22, where panel 9 (1.88) and panel 8 (2.07) show the smallest average absolute differences, whereas panel 18 (9.22) and panel 16 (6.63) show the largest average absolute differences.

For intermediate education, 12 panels reveal lower point estimates than the face-to-face probability sample PSLC (up to -13.38 for panel 18), whereas 4 report large point estimates than the benchmark (up to 8.25 for panel 6). For full employment, 15 panels report smaller point estimates compared to face-to-

face probability sample LFS (up to -13.00 for panel 14), whereas 3 panels report larger point estimates compared to the face-to-face probability sample LFS, which are very large for panel 3 (15.50) and panel 18 (17.78).

4.1.4 Non-register variables – HT estimator LISS

Comparing HT estimates for education and employment only, LISS has an average absolute difference of .86 compared to GREG estimates of LFS or PSLC. Looking at the individual variables, we see that for education, 16 of the 18 non-probability Internet panels differ significantly from the face-to-face probability sample compared to LISS. For employment, 14 of the 18 panels are significantly different from the face-to-face probability sample compared to LISS.

In summary, the results reveal that panel 8 and 9 are the more accurate online non-probability panels for the non-register variables, whereas panel 18, panel 3, but also panel 13 and 16 perform rather poorly since they show large average absolute differences and strongly significant differences for the individual variables. LISS performs well since for 83% of the variables the differences of NOPVO compared to the benchmark is significantly larger than for LISS.

4.1.5 Non-register variables – GREG estimator NOPVO.

With the non-register variables the question is addressed if weighting with auxiliary variables, available from registers improves the accuracy of the NOPVO panels. Therefore GREG estimates of the non-register variables for LISS and NOPVO are compared with the benchmarks, i.e. GREG estimates of the two probability samples. Results are summarized in Table 5 and Figure 1 in the appendix.

Using the two non-register variables measured comparably across all surveys (education and employment) after weighting, the non-probability Internet samples show average absolute difference to the face-to-face probability samples ranging from 1.94 for panel 2 and 2.13 for panel 8 to 14.45 for panel 18 (see Table 5). If we consider all four non-register variables, the average absolute differences of the non-probability Internet panels range from 2.11 to 8.60. Panel 9 shows the smallest average difference with 2.11, followed by panel 7 (2.61). Panel 18 again performs worst (8.60), followed by panel 11 (6.65).

Weighting reduced the deviation from the probability face-to-face samples for only 5 of the 18 non-probability Internet samples (panel 3, 7, 14, 16 and 18 when looking at employment and education only, panel 3, 6, 10, 16 and 18 when looking at all 4 non-register variables). Panel 6 shows the largest increase in the average absolute difference after weighting (for education and employment only) with an increase of 3.26 (from 7.97 to 11.23). Most interestingly, all of the non-probability Internet samples, also

panel 8, showed larger average absolute differences to the face-to-face probability sample than LISS after weighting. This result indicates that weighting does not improve non-probability samples to the level of a probability sample.

For education, 11 panels show point estimates that are lower than PSLC. The largest deviation is with -15.01 (panel 18) now even higher than for the non-weighted results. 7 panels show higher point estimates, and again the largest positive difference (10.71 for panel 6) is higher than for the non-weighted results. For employment, 15 panels show lower point estimates than the face-to-face probability sample LFS. The largest difference can be observed for panel 10 (12.80), which is even 2.05 percentage points higher than for the non-weighted results. 3 panels report higher point estimates for full employment, where the difference for panel 18 decreased by 3.89 to 13.89, but the large difference for panel 3 remains more or less the same (15.88 versus 15.50 in the previous analyses).

4.1.6 Non-register variables – GREG estimator LISS.

The average absolute deviation from LISS compared to the face-to-face probability samples was .99, which is smaller than all non-probability panels but higher compared to the unweighted results (0.86). Weighting resulted in a smaller difference for education, but weighting employment was less successful. The difference to LFS is now 1.37 compared to 0.98 for the non-weighted results. If we look at the individual variables, we see that for education, 16 of the 18 non-probability Internet panels differ significantly to the face-to-face probability sample PSLC compared to LISS. For employment, 17 of the 18 panels differ significantly from the face-to-face probability sample LFS compared to LISS.

In summary, for 92% of the variables the differences of NOPVO compared to the benchmark is significantly larger than for LISS if the GREG estimator is applied. That is an increase of 9% compared to the HT estimator. This illustrates that weighting with respect to auxiliary variables does not increase the precision of the NOPVO panels.

Two additional observations raise concerns about the accuracy of the non-probability panels when looking at the ranking of the panels according to the absolute magnitude of the differences. First, the non-probability panels often differ substantially from one another in their performance for the register and non-register variable estimates. For example, panel 16 was the most accurate non-probability panel for the register variables (3/21), but ranks 17 out of 18 for the unweighted non-register variables and 14 out of 18 for the weighted register variables. In contrast, panel 9 ranks 17/21 for the register variables, but is the most accurate non-probability panel along the non-register variables. The correlation between the ranks of the firms in terms of accuracy with the register and the non-register variables, both unweighted and weighted, the correlations are tiny (ranging from -.06 to -.10).

Second, the rank position of the non-probability panels is affected by weighting, and by the variables that are included. Panel 7, for example, improves 5 ranks after weighting if we look at employment and education only (from rank 9 to rank 4) or 6 ranks (from 8 to 2) if we look at all register variables. Panel 18 ranks equally low before and after weighting (rank 19 for education and employment only and 18 for all non-register variables). Panel 3 improves 5 ranks after weighting (from 16-11) if we look at all non-register variables, but only 1 rank if we consider education and employment only. Panel 6 drops from rank 14 to 18 after weighting if we look at education and employment only, whereas it drops only from rank 13 to 15 if we look at all non-register variables. These findings demonstrate the lack of consistency in the performance of the non-probability samples.

4.2 Relations Between Variables

To explore relations between variables, we regressed life satisfaction on known predictors of it: health, age, and gender (e.g., Daig et al. 2009; Lang et al. 2013). The null hypothesis that the OLS estimates for the regression coefficients are equal for the eighteen Internet non-probability panels[4] and the probability face-to-face sample was tested against the alternative hypothesis that at least one pair of surveys yielded coefficients that were significantly different from one another using a partial F-test. In the unrestricted regression, we allow for panel specific intercepts and panel specific regression coefficients for the auxiliary variables to account for differences between panels and isolate the effects of life satisfaction on our variable of interest. We also gauged the extent to which selection bias in the estimates of the regression coefficients can be removed calculating weighted LS estimates for the regression coefficients using the GREG weights.

4.2.1 Regressions – unweighted.

When life satisfaction was regressed on health, the restricted and unrestricted models were significantly different from one another ($p < .001$). The regression coefficients for the interaction terms range from -.08 (panel 3) to .16 (panel 15). (see grey area in Table 6 and Figure 2 in the appendix). When adding up the main effect for health, the unique effects per panel as captured by the panel dummies, and the respective interaction effect, one would come to different conclusions about the strength of the relationship of life satisfaction on health. For example, panel 5 would suggest an effect of .36 versus .68 for panel 11. Panel 18 would predict an effect (.42) which is very close to PSLC (.43).

When regressing life satisfaction on gender, the partial F-test indicates again that the regression coefficients for the different panels are significantly different ($p < .001$). According to PSLC, life satisfaction is not related to gender ($\beta = .01$; p-value = .69); The regression coefficients for the

interactions ranged from -.21 (panel 11) to .25 (panel 16). When adding up the main effect for gender, the unique effects per panel as captured by the panel dummies, and the respective interaction effect, the results differ where panel 2 and 18 suggest very similar effects to PSLC (.08 and .09 respectively), but panel 16 suggests that gender has a rather strong effect of .69 on life satisfaction. Also when regressing life satisfaction on age, the partial F-test reject the hypothesis that the regression coefficients for the different panels are equal ($p < .001$). The regression coefficients for the interactions ranged from -.36 (panel 17) to .15 (panel 18). When adding up the main effect for age, the unique effects per panel as captured by the panel dummies, and the respective interaction effect, the results differ rather drastically where panel 17 suggests a negative relationship of life satisfaction on age (-.16) but panel 16 suggests a positive relationship (.56).

The results show that for all three variables, researchers would come to rather different conclusions about the strength of relationships and in the case of age even the sign of the relationship between variables.

4.2.2 Regressions – weighted.

The results for the regression analyses using weighted data are also strikingly similar to the unweighted results (see Table 6 and Figure 2). As with the unweighted data, the partial F-test indicates that the regression coefficients for life satisfaction on health in the different panels are significantly different (*p*-value < .001), with interaction coefficients ranging from -.06 (panel 13) to .15 (panel 12). When adding up the main effect for age, the unique effects per panel as captured by the panel dummies, and the respective interaction effect, the effects range from .37 (panel 5) to .70 (panel 14, where panel 12 and 17 predict similar results (.42) than PSLC (.44). Also for life satisfaction on gender, the F-test reject the hypothesis that the regression coefficients for the different panels are equal ($p < .001$), with coefficients ranging from -.19 (panel 2) to -.41 (panel 16). The total effects range from .01 (panel 2 to .74 (panel 16), where panel 2 and PSLC report the same, insignificant coefficient of .01. A similar result is obtained for the regression coefficients of life satisfaction on age where the restricted and unrestricted models were again significantly different from one another ($p < .001$), with interaction coefficients ranging from -.44 (panel 18) to .15 (panel 4). The total effects differ again quite significantly. Panel 10 (.02) and 7 (.04) come to similar results than PSLC (.01), but panel 17 suggests a negative relationship of life satisfaction on age of -.27, whereas panel 14 reports an effect of .45. Again, researchers would not only come to different conclusions about the strength of relationships between variables, but even about the sign of a relationship. In summary it can be concluded that a weighted least squares regression does not solve the problems with significantly different relations between variables in different panels.

4.3 Panel Management Techniques

Spearman correlations of panel management techniques with the different accuracy metrics reveal only very few significant correlations, which may also be due to the small number of panels. Nevertheless, the results provide a few interesting insights (see Table 7 in the appendix for details). With respect to panel management, we find that panel size has a weak positive correlation with the average difference to the benchmark for register variables. Older panels are somewhat correlated with lower average differences for register as well as non-register variables. Moreover, the practice of panel refreshment is somewhat correlated to smaller average deviations from the benchmarks for both register and non-register. Surprisingly, recruiting panel members through links on websites is correlated with lower differences for non-register variables, and using snowball methods is correlated with lower differences to the benchmark for register variables. The more traditional telephone recruitment is correlated with lower deviations for non-register variables, and recruiting from existing mail panels is somewhat correlated with lower differences to the benchmarks for register variables. Panels who buy addresses should be avoided since this practice is somewhat correlated with higher deviations for both register and non-register variables. With respect to the incentives we find that prizes through lotteries are negatively correlated with differences to the benchmark for register variables. Using lottery tickets is somewhat correlated with larger deviations from the benchmark for register variables, but correlated with lower differences from the benchmarks for non-register variables. Using rewards for registering in a panel is somewhat correlated with larger deviations from the benchmark for register variables, but significantly correlated with lower differences from the benchmarks for non-register variables.

5. Discussion

A AAPOR (American Association for Public Opinion Research) task force concludes that "increasing nonresponse in traditional methods, rising costs and shrinking budgets, dramatic increases in Internet penetration, the opportunities in questionnaire design on the Web, and the lower cost and shorter cycle times of online surveys—continue to increase the pressure on all segments of the survey industry to adopt online research methods" (Baker et al. 2010). However, this approach is overwhelmingly focused on maximizing the amount of data that can be obtained for a given budget, thereby ignoring the risk of introducing large amounts of selection bias. Modern calibration techniques, known from survey sampling to correct for selective nonresponse, are considered to correct for this selection bias. Our results challenge the frequent suggestion from proponents of non-probability Internet survey sampling that it may be possible to correct for the sort of discrepancies observed in this analysis by adjusting the data to more closely resemble the population of interest. We find that weighting does not reduce selection bias in the level estimates and relationships between variables of non-probability online panels. The available auxiliary information that is used to construct weighting models often fail to explain the non-random selection process that led to the realized panel compositions.

Our findings also draw attention to the inconsistent performance of online non-probability panels. Panels that perform well for the register variable estimates do not necessarily perform well for non-register estimates. Also, across the non-probability Internet surveys, some variables are estimated incorrectly by only a subset of panels (e.g. gender), whereas other variables (e.g. education or employment) are estimated incorrectly by most panels. Some panels simply perform poorly for most variables.

Moreover, studying the *relationship* between variables, which is very relevant in (social) science in general, led to the conclusion that panels differ significantly from each other with respect to the strength and even sign of relationships. The results do not improve when weighting the data. Differences between the online non-probability panels and the Statistic Netherlands probability sample are even more pronounced if with a weighted least squares approach. In other words, researchers may draw substantively different conclusions regarding the strength and sign of relationships between variables, depending on their choice of online panel. For survey researchers, this implies that choosing an online panel is extremely difficult since the performance varies depending on the type of variable (register vs. non-register) or data (weighted vs. unweighted) that is used.

The first implication of our research is that researchers should remain true to the traditional concept of random sampling to select samples that are representative for the intended target population. As Brick (2011) notes in is contribution to the 75th anniversary issue of Public Opinion Quarterly, the

development of online panels is driven by practical requirements for timely and cost-efficient survey research rather than by statistics theory. A common refrain from the commercial survey industry is therefore that “a probability sample with a low response rate or coverage ratio is no “better” than a nonprobability or volunteer sample” (Brick 2011). Yet, our results fundamentally challenge the quality of nonprobability panels and suggest sticking to the foundations of survey sampling. The most straightforward way to circumvent the problems that we identified is to apply the traditional concept of random sampling to select samples that are representative for the intended target population and make an all-out effort to achieve survey participation for all sampled units. Bethlehem and Biffignandi (2011) provide a comprehensive overview of methods to obtain representative web panels. The LISS panel that is based on this concept and in addition pays a lot of effort to maintain a representative panel, indeed has better accuracy measures than the 18 online non-probability panels in the Netherlands. But our data also reveals that some panel management techniques correlated with accuracy measures. For example, researchers interested in register variables should maybe avoid larger panels, panels that buy addresses, and panels that use lottery tickets for lotteries, point systems, and rewards for registering in the panel since our results reveal some correlations with higher differences to the benchmarks. In contrast, recruiting from existing mail panels, panel refreshment and prizes through lotteries or monetary incentives are somewhat negatively correlated with differences to the benchmark. Surprisingly, also recruitment through snowball techniques shows a large and significant negative correlation with average differences to the benchmarks for the register variables.

For non-register variables older panels and panel refreshment also perform slightly better and show some negative correlations to the average difference to the benchmarks. Furthermore, recruiting through telephone, lottery tickets for lottery, and using rewards for registration are correlated with lower differences to benchmarks. Surprisingly, recruiting panel members through links on websites is correlated with lower differences to the benchmark for non-register variables. Overall, it should be said however that the correlations are small and only few are significant, and the results are not always consistent across register and non-register variables and panels.

A second implication is that marketers need to pay more attention to the sampling procedure and quality of data. However, this information is oftentimes not provided. Table 7 provides an overview of all articles published in the Journal of Marketing and Journal of Marketing Research between 2009-2013 which conduct surveys. Although information on the sampling frame and mode is provided for most studies, only 36% (JM) and 41% (JMR) provide information on the sampling method. The fact that 36% (JM) and 22% (JMR) of all survey data are collected on the Internet raises the question whether some of the studies are conducted through online non-probability samples without providing proper information on the sampling method and potential biases that may affect the results. In the worst case, many

samples may be woefully inadequate and provide biased results despite favorable psychometric properties and sophisticated analyses.

Our results also challenge the relevance of response rates for non-probability panels, which many still consider to be the key survey metric. The accuracy measures for the different non-probability panel do not correspond to their sample sizes, which challenges the traditional wisdom that response rates provide us with the indication of nonresponse error. We like to call on the marketing community to provide more elaborate and honest account of the nature of a sample and its potential biases. To quote Churchill (Churchill 1979, p. 73) we should "reduce the prevalent tendency to apply extremely sophisticated analysis to faulty data and thereby execute still another GIGO [garbage in, garbage out] routine."…As scientists, marketers should be willing to make this commitment to "quality research".

Although the benchmarks for the register variables from the municipal basis administration are very strong since they contain no sampling error or selection bias, the benchmarks for the non-register variables come from two high quality, face-to-face probability government surveys. We cannot completely rule out context or mode effects, although most variables that are included in our study are not sensitive for measurement biases. It would have been better if all samples (LFS, PSLC and NOPVA) had used the same questionnaire. Since the pattern of our results for the register and non-register variables is rather similar we have confidence in our results, but future research should generalize our findings with different benchmarks.

In this paper we proposed a test to compare the deviation of the modal response category of two unequally sized samples from a benchmark. Of course it would also be interesting to analyze deviations from a benchmark across all response categories simultaneously, but then future research first has to develop a generalized version of our significance test using a multinomial distribution.

Future research should also generate more empirical evidence. It would be especially interesting to conduct a similar study in Scandinavian countries since they do not only have very high Internet adoption rates as well (Seybert 2012), but they also have very reliable municipal registers (United Nations Economic Commission for Europe 2007). It would also be interesting to repeat this study nowadays although. Albeit higher Internet penetration in 2016 compared to 2006 (http://www.internetworldstats.com/top25.htm) ,response rates to surveys have decreased strongly over the years (http://www.nsf.gov/news/special_reports/survey/index.jsp?id=question), so that we would not expect the online non-probability panels to perform better nowadays than in our dataset. In addition, the variables selected for comparison could be extended to continuous variables to see if they show similar results.

References

Baker, R., S. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, M. Dennis, D. Dillman, M. R. Frankel, P. Garland, R. M. Groves, C. Kennedy, J. A. Krosnick, D. Lee, P. J. Lavrakas, M. Link, L. Piekarski, K. Rao, R. K. Thomas, and D. Zahs (2010), "AAPOR Report on Online Panels," Public Opinion Quarterly, 74 (4), 711-81.

Bakker, B. F. M. (2012). Estimating the validity of administrative variables. *Statisitica Neerlndica,* 66 (1), 8-17.

Bethlehem, Jelke and Silvia Biffignandi (2011), Handbook of Web Surveys. New-York: Wiley.

Brick, J. Michael (2011), "The Future of Survey Sampling," Public Opinion Quarterly, 75 (5), 872-88.

Chang, Linchiat and Jon. A. Krosnick (2009), "National Surveys Via RDD Telephone Interviewing Versus The Internet: Comparing Sample Representativeness and Response Quality," Public Opinion Quarterly, 73 (4), 641-78.

Churchill, Gilbert A. Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," Journal of Marketing Research, 16 (1), 64-73.

Couper, Mick P. (2000), "Web Surveys: A Review of Issues and Approaches," Public Opinion Quarterly, 64 (4), 464-94.

Daig, Isolde, Peter Herschbach, Anja Lehmann, Nina Knoll, and Oliver Decker (2009), "Gender and Age Differences in Domain-Specific Life Satisfaction and the Impact of Depressive and Anxiety Symptoms: A General Population Survey from Germany," Quality of Life Research, 18 (6), 669-78.

Dever, Jill A., Ann Rafferty, and Richard Valliant (2008), "Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?," Survey Research Methods, 2 (2), 47-62.

Frölich, Markus (2004), "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators," Review of Economics & Statistics, 86 (1), 77-90.

Heckman, James J., Hidehiko Ichimura, and Petra Todd (1998), "Matching as an Econometric Evaluation Estimator," Review of Economic Studies, 65 (2), 261-94.

Horvitz, Daniel G. and Donovan J. Thompson (1952), "A Generalization of Sampling Without Replacement from a Finite Universe," Journal of the American Statistical Association, 47 (260), 663-85.

http://www.harrisinteractive.com/partner/methodology.asp.

http://www.rflonline.com/clientsummit/notes/Kim-Dedeker-notes.html.

Huang, E.T. and W.A. Fuller (1978), "Nonnegative Regression Estimation for Survey Data," in Proceedings of the Section on Social Statistics, . Alexandria: American Statistical Association.

Ichimura, Hidehiko and Christopher Taber (2001), "Propensity-Score Matching with Instrumental Variables," American Economic Review, 91 (2), 119-24.

Inside Research (2009), "U.S. Online MR Gains Drop," 20, 1 (11-134).

Kivlin, Joseph E. (1965), "Contributions to the Study of Mail-Back Bias," Rural Sociology, 30 (3), 322-26.

Lang, Frieder R. , David Weiss, Denis Gerstorf, and Gert G. Wagner (2013), "Forecasting Life Satisfaction Across Adulthood: Benefits of Seeing a Dark Future?," Psychology and Aging, 28 (1), 249-61.

Lee, Sunghee (2006), "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys," Journal of Official Statistics, 22 (2), 329-49.

Malhotra, Neil and Jon. A. Krosnick (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 Nonprobability Samples," Political Analysis, 15 (3), 286-323.

Postoaca, Andrei (2006), The Anonymous Elect. Berlin, Germany: Springer.

Särndal, Carl-Erik, Bengt Swensson, and Jan Wretman (1992), Model Assisted Survey Sampling. New-York: Springer-Verlag.

Scherpenzeel, Annette (2009), "Start of the LISS Panel: Sample and Recruitment of a Probability-Based Internet Panel," in CentERdata Vol. 2012.

Seybert, Heidi (2012), "Internet Use in Households and by Individuals In 2012: One Third Of Europeans Used the Internet on Mobile Devices AwayfFrom Home or Work," EUROSTAT: Statostics in Focus, 50/2012.

United Nations Economic Commission for Europe (2007), Register-Based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics New York and Geneva: United Nations.

Valliant, Richard and Jill A. Dever (2011), "Estimating Propensity Adjustments for Volunteer Web Surveys," Sociological Methods & Research, 40 (1), 105-37.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang (2011), "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples," Public Opinion Quarterly, 75 (4), 709-47.

Footnotes

[1] Of course, sampling theory only anticipates increased accuracy with increased sample size if the sample is drawn randomly from the population, so this logic does not apply to the non-probability samples.

[2] In fewer than 5% of the households, data about one or more residents were not collected directly or via a proxy, and these households were treated as not responding and were dropped from the analyses.

[3] For education and employment, it was necessary to combine response choices with one another to produce comparability. The full employment category differs marginally between the three data sources. See Appendices 1and 2 for more details on the exact wordings.

[4] LISS was not included in this analysis, since their question measuring life satisfaction was worded different from that in the PSLC and the eighteen non-probability sample panels.

Appendix for Tables and Figures

Table 1: Panel Management Techniques Used by Non-Probability Internet Survey Firms

						Panel recruitment							Incentives				
	Founding year	Number of active panel members	Number of invitations per month	Percentage drop-out of active panel members per year	Panel refresh	Links on websites	From research conduced through traditional modes	Bought addresses	Snowball method	Telephone recruitment	Existing mail panels	Other	Use of reward for registering in panel	Prizes through lottery	Monetary incentive	Lottery tickets for lottery	Point system
1	2001	15000	unlimited	10%	No	21%	13%	20%	14%	10%	12%	10%	No	1	0	0	1
4	2000	15000	2	2%	Yes	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	0	0	0	1
5	2001	25000	2	2%	Yes	50%	10%	29%	1%	10%	0%	0%	Yes	0	0	0	1
6	2004	24331	2	10%	Yes	12%	0%	79%	0%	9%	0%	0%	No	1	0	0	1
7	2000	46178	unlimited	11%	Yes	32%	28%	0%	3%	16%	0%	21%	Yes	0	0	0	1
8	2002	57000	2	8%	yes	68%	0%	22%	0%	0%	4%	6%	Yes	0	0	1	1
9	before 2000	106000	1	8%	No	90%	5%	2%	2%	1%	0%	0%	No	1	0	1	1
10	before 2000	6500	1	15%	No	10%	10%	50%	10%	0%	0%	20%	No	1	0	0	0
11	2000	127000	unlimited	16%	Yes	0%	0%	0%	0%	0%	0%	100%	No	1	1	0	0
12	before 2000	83534	1	10%	Yes	10%	5%	0%	5%	0%	5%	75%	No	0	1	0	1
13	before 2000	93000	unlimited	20%	Yes	0%	36%	0%	22%	27%	0%	15%	No	1	1	0	0
14	2002	226000	1	30%	No	25%	0%	0%	25%	0%	0%	50%	No	0	0	0	1
15	2002	97400	4	2%	Yes	80%	0%	0%	10%	0%	0%	10%	No	1	0	0	1
16	2005	16872	1	4%	Yes	60%	30%	0%	5%	5%	0%	0%	No	1	0	0	1
17	2004	17955	1	26%	No	0%	95%	0%	5%	0%	0%	0%	No	0	0	0	1
18	before 2000	148805	2	10%	No	0%	70%	20%	10%	0%	0%	0%	No	0	0	0	1
Total				**12%**		**31%**	**20%**	**15%**	**7%**	**5%**	**1%**			**8**	**3**	**2**	**14**

**Panel 2 and 3 did not provide information on their panel management*

Table 2: Overview of variables used to construct weights for each online panel

Sample	Variables used for weighting	# of Negative Weights
Non-probability Internet sample 1	Age (5) * Gender (2) + People in HH (6) + Urbanization (5) + Province (12) + Nationality (3)	1
Non-probability Internet sample 2	People in HH (6) + Gender (2) + Urbanization (5) + Province (12) + Nationality (3)	5
Non-probability Internet sample 3	Province (12) + Age (5) * Gender (2) + People in HH (6)	0
Non-probability Internet sample 4	Age (5) * Gender (2) + Urbanization (6) + Province (12) + People in HH (6)	0
Non-probability Internet sample 5	Age (5) * Gender (2) + Urbanization (5) + Province (12) + People in HH (6) + Nationality (3)	0
Non-probability Internet sample 6	Age (5) * Gender (2) + Urbanization (5) + Province (12) + People in HH (6) + Nationality (3)	17
Non-probability Internet sample 7	Age (5) * Gender (2)	0
Non-probability Internet sample 8	Province (12) + Age (5) * Gender (2) + Nationality (5) + People in HH (6)	0
Non-probability Internet sample 9	Age (5) * Gender (2) + Urbanization (6) + Province (12)	0
Non-probability Internet sample 10	Province (12) + Age (5) * Gender (2) + Nationality (3) + People in HH (6)	8
Non-probability Internet sample 11	Age (5) * Gender (2) + Urbanization (5) + Province (12) + People in HH (6)	0
Non-probability Internet sample 12	Age (5) * Gender (2) + Urbanization (5) + People in HH (6)	0
Non-probability Internet sample 13	Age (5) * Gender (3) + Urbanization (5) + Province (12) + People in HH (6)	1
Non-probability Internet sample 14	Province (12) + Age (5) * Gender (3) + People in HH (6)	0
Probability Internet sample	People in HH (6) + Age (5) * Gender (2) + Urbanization (5) + Province (12)	
Probability Face-to-Face Sample 1 (PSLC)	Month (12) + Gender (2)*age (3) *marital status (2) + Gender (2)*age (17) + Marital status (4) + Housholdsize (5) + Urbanisation level (5) + Province plus 4 bigest citys (16) + Age (3)*region (4)	
Probability Face-to-Face Sample 1 (LFS)	Household type (3) + Income(6) + Urbanisation level (5) + [Registered as being unemployed * province(12) + not registered being unemployed] + [Duration registered unemployed (5) + not registered being unemployed] + Gender (2)* Nationality (10)+ Gender (2)*age (38)	

The numbers in the parantheses refer to the number of categories; see Appendix A for more information

Table 3: Accuracy of Register Variables, HT estimator

	Benchmark		Probability Panels			Non-Probability Panels																
Data Source	**Municipality Register**		**LFS**	**PSLC**	**LISS**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Year	**2006**	**2008**	**2006**	**2006**	**2008**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**
N			**10589**	**9607**	**5172**	**769**	**621**	**536**	**603**	**548**	**654**	**675**	**385**	**184**	**403**	**488**	**562**	**695**	**408**	**191**	**388**	**388**	**604**
Gender: Male	50.42	50.34																					
Point Estimate			49.27	49.10	48.90	48.50	46.70	28.73	49.25	38.50	54.28	48.89	54.29	44.02	49.88	52.46	48.93	48.20	36.76	48.69	54.90	47.16	48.51
Δ Benchmark			-1.16	-1.33	-1.44	-1.92	-3.72	-21.69	-1.17	-11.92	3.86	-1.53	3.87	-6.40	-0.54	2.04	-1.49	-2.22	-13.66	-1.73	4.48	-3.26	-1.91
p-value: difference to benchmark compared to LFS					*0.03*	*0.04*	*0.00*	*0.00*	*0.40*	*0.00*	*0.00*	*0.17*	*0.00*	*0.00*	*0.85*	*0.06*	*0.21*	*0.01*	*0.00*	*0.18*	*0.00*	*0.00*	*0.05*
p-value: difference to benchmark compared to PSLC					*0.24*	*0.09*	*0.00*	*0.00*	*0.55*	*0.00*	*0.00*	*0.29*	*0.00*	*0.00*	*0.90*	*0.12*	*0.33*	*0.03*	*0.00*	*0.25*	*0.00*	*0.00*	*0.11*
p-value: difference to benchmark compared to LISS						*0.15*	*0.00*	*0.00*	*0.64*	*0.00*	*0.00*	*0.39*	*0.00*	*0.00*	*0.93*	*0.17*	*0.42*	*0.05*	*0.00*	*0.30*	*0.00*	*0.00*	*0.17*
Age: 35-44	24.81	24.32																					
Point Estimate			24.28	25.27	25.10	27.57	33.98	33.21	25.21	30.47	23.24	22.96	25.71	28.26	24.32	23.77	22.78	24.17	34.56	22.51	23.97	23.20	25.99
Δ Benchmark			-0.53	0.46	0.78	2.76	9.17	8.40	0.40	5.66	-1.57	-1.85	0.90	3.45	-0.49	-1.04	-2.03	-0.64	9.75	-2.30	-0.84	-1.61	1.18
p-value: difference to benchmark compared to LFS					*0.00*	*0.00*	*0.00*	*0.00*	*0.04*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.21*	*0.00*	*0.00*	*0.06*	*0.00*	*0.00*	*0.05*	*0.00*	*0.00*
p-value: difference to benchmark compared to PSLC					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.07*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
p-value: difference to benchmark compared to LISS						*0.00*	*0.00*	*0.00*	*0.30*	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*	*0.55*	*0.08*	*0.00*	*0.40*	*0.00*	*0.00*	*0.27*	*0.01*	*0.00*
Origin of self: Netherlands	89.40	89.30																					
Point Estimate			89.99	89.65	93.91	96.62	97.42	97.01	97.01	95.80	95.87	97.63	96.36	96.74	95.53	95.90	95.73	97.41	93.63	95.81	92.27	97.16	96.36
Δ Benchmark			0.59	0.25	4.61	7.22	8.02	7.61	7.61	6.40	6.47	8.23	6.96	7.34	6.13	6.50	6.33	8.01	4.23	6.41	2.87	7.76	6.96
p-value: difference to benchmark compared to LFS					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
p-value: difference to benchmark compared to PSLC					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
p-value: difference to benchmark compared to LISS						*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.01*	*0.00*	*0.02*	*0.05*	*0.08*	*0.03*	*0.03*	*0.00*	*0.61*	*0.11*	*0.95*	*0.00*	*0.00*
Origin of mother: Netherlands	82.51	82.61																					
Point Estimate			88.26	85.96	91.07	93.89	95.97	94.59	93.86	93.98	93.12	95.26	91.95	92.93	91.56	94.67	91.64	93.67	89.71	92.67	88.66	95.10	94.87
Δ Benchmark			5.75	3.45	8.46	11.38	13.46	12.08	11.35	11.47	10.61	12.75	9.44	10.42	9.05	12.16	9.13	11.16	7.20	10.16	6.15	12.59	12.36
p-value: difference to benchmark compared to LFS					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.10*	*0.01*	*0.39*	*0.00*	*0.00*
p-value: difference to benchmark compared to PSLC					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
p-value: difference to benchmark compared to LISS						*0.00*	*0.00*	*0.00*	*0.01*	*0.01*	*0.03*	*0.00*	*0.23*	*0.15*	*0.33*	*0.00*	*0.27*	*0.01*	*0.81*	*0.19*	*0.96*	*0.00*	*0.00*
Origin of father: Netherlands	82.87	82.97																					
Point Estimate			88.31	86.40	91.69	94.02	96.14	96.27	93.37	94.16	92.97	93.33	92.47	94.02	93.30	93.85	90.57	94.96	90.69	91.10	86.60	96.13	94.87
Δ Benchmark			5.44	3.53	8.72	11.15	13.27	13.40	10.50	11.29	10.10	10.46	9.60	11.15	10.43	10.98	7.70	12.09	7.82	8.23	3.73	13.26	12.00
p-value: difference to benchmark compared to LFS					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.03*	*0.06*	*0.94*	*0.00*	*0.00*
p-value: difference to benchmark compared to PSLC					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.40*	*0.00*	*0.00*
p-value: difference to benchmark compared to LISS						*0.01*	*0.00*	*0.00*	*0.06*	*0.02*	*0.10*	*0.06*	*0.29*	*0.12*	*0.10*	*0.04*	*0.79*	*0.00*	*0.76*	*0.60*	*1.00*	*0.00*	*0.00*

Data Source	Municipality Register		LFS	PSLC	LISS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Year	**2006**	**2008**	**2006**	**2006**	**2008**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**
People in household: 2	30.80	30.50																					
Point Estimate			33.66	33.12	33.85	31.34	23.51	33.02	N/A	33.03	32.26	34.81	34.81	38.04	34.58	34.02	28.29	34.10	28.15	30.53	37.11	32.28	35.10
Δ Benchmark			2.86	2.32	3.35	0.54	-7.29	2.22		2.23	1.46	4.01	4.01	7.24	3.78	3.22	-2.51	3.30	-2.65	-0.27	6.31	1.48	4.30
p-value: difference to benchmark compared to LFS					*0.02*	*1.00*	*0.00*	*0.84*		*0.79*	*0.99*	*0.03*	*0.09*	*0.00*	*0.12*	*0.32*	*0.64*	*0.27*	*0.62*	*1.00*	*0.00*	*0.97*	*0.03*
p-value: difference to benchmark compared to PSLC					*0.00*	*1.00*	*0.00*	*0.59*		*0.51*	*0.94*	*0.00*	*0.02*	*0.00*	*0.03*	*0.11*	*0.33*	*0.06*	*0.35*	*0.99*	*0.00*	*0.89*	*0.00*
p-value: difference to benchmark compared to LISS						*1.00*	*0.00*	*0.95*		*0.92*	*1.00*	*0.15*	*0.22*	*0.00*	*0.28*	*0.57*	*0.84*	*0.55*	*0.80*	*1.00*	*0.00*	*0.99*	*0.12*
Region: West	30.22	30.33																					
Point Estimate			32.79	34.12	27.83	26.27	25.60	25.75	19.73	24.13	25.88	27.56	27.27	25.54	22.58	24.18	23.49	25.18	25.00	31.94	26.03	24.74	24.67
Δ Benchmark			2.57	3.91	-2.50	-3.95	-4.62	-4.47	-10.49	-6.09	-4.34	-2.66	-2.95	-4.68	-7.64	-6.04	-6.73	-5.04	-5.22	1.72	-4.19	-5.48	-5.55
p-value: difference to benchmark compared to LFS					*0.62*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.47*	*0.29*	*0.05*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.73*	*0.02*	*0.00*	*0.00*
p-value: difference to benchmark compared to PSLC					*1.00*	*0.45*	*0.19*	*0.28*	*0.00*	*0.01*	*0.27*	*0.97*	*0.82*	*0.29*	*0.00*	*0.01*	*0.00*	*0.06*	*0.08*	*0.94*	*0.35*	*0.05*	*0.02*
p-value: difference to benchmark compared to LISS						*0.01*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.42*	*0.26*	*0.04*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.71*	*0.02*	*0.00*	*0.00*
Province: South Holland	21.24	21.20																					
Point Estimate			19.59	20.49	18.55	22.76	23.19	13.11	21.72	21.21	23.55	22.96	N/A	25.00	21.59	19.88	22.95	3.74	19.12	25.13	19.85	21.96	19.37
Δ Benchmark			-1.65	-0.75	-2.65	1.52	1.95	-8.13	0.48	-0.03	2.31	1.72	N/A	3.76	0.35	-1.36	1.71	-17.50	-2.12	3.89	-1.39	0.72	-1.87
p-value: difference to benchmark compared to LFS					*0.00*	*0.61*	*0.23*	*0.00*	*1.00*	*1.00*	*0.08*	*0.44*		*0.03*	*0.99*	*0.69*	*0.45*	*0.00*	*0.23*	*0.01*	*0.62*	*0.95*	*0.30*
p-value: difference to benchmark compared to PSLC					*0.00*	*0.01*	*0.00*	*0.00*	*0.83*	*0.98*	*0.00*	*0.01*		*0.00*	*0.80*	*0.08*	*0.01*	*0.00*	*0.00*	*0.00*	*0.07*	*0.56*	*0.00*
p-value: difference to benchmark compared to LISS						*0.98*	*0.84*	*0.00*	*1.00*	*1.00*	*0.66*	*0.95*		*0.22*	*1.00*	*0.97*	*0.93*	*0.00*	*0.76*	*0.14*	*0.95*	*1.00*	*0.88*
Average absolute Δ benchmark			**2.57**	**2.00**	**4.07**	**5.06**	**7.69**	**9.75**	**6.00**	**6.89**	**5.09**	**5.40**	**5.39**	**6.81**	**4.80**	**5.42**	**4.70**	**7.49**	**6.58**	**4.34**	**3.74**	**5.77**	**5.77**
Comparison to benchmark compared to LFS:																							
p-value					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.06*	*0.08*	*0.00*	*0.00*
Fraction of significant statistical tests					*88%*	*75%*	*88%*	*88%*	*71%*	*75%*	*75%*	*63%*	*71%*	*100%*	*50%*	*63%*	*63%*	*75%*	*63%*	*50%*	*63%*	*75%*	*88%*
Comparison to benchmark compared to PSLC:																							
p-value					*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.01*	*0.00*	*0.00*
Fraction of significant statistical tests					*75%*	*63%*	*88%*	*75%*	*71%*	*75%*	*75%*	*75%*	*86%*	*88%*	*63%*	*63%*	*75%*	*75%*	*75%*	*63%*	*63%*	*75%*	*88%*
Comparison to benchmark compared to LISS:																							
p-value					*0.01*	*0.10*	*0.00*	*0.00*	*0.01*	*0.00*	*0.09*	*0.04*	*0.11*	*0.04*	*0.21*	*0.07*	*0.22*	*0.00*	*0.01*	*0.38*	*0.62*	*0.05*	*0.03*
Fraction of significant statistical tests						*63%*	*88%*	*88%*	*43%*	*75%*	*63%*	*38%*	*43%*	*63%*	*13%*	*50%*	*38%*	*75%*	*38%*	*13%*	*38%*	*75%*	*63%*
Rank by average Δ benchmark			**2**	**1**	**4**	**8**	**20**	**21**	**15**	**18**	**9**	**11**	**10**	**17**	**7**	**12**	**6**	**19**	**16**	**5**	**3**	**14**	**13**

Table 4: Accuracy of Non-Register Variables, HT estimator

	Benchmarks			Probability Panel	Non-Probability Panels																
Data Source				**LISS**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Year	**2006**	**2008**		**2008**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**
Education: Intermediate																					
Point Estimate	36.84	37.31	a	37.66	32.15	37.82	33.66	39.70	40.97	45.09	30.94	36.80	40.33	33.25	31.13	37.32	30.60	26.36	34.24	29.87	28.72	23.46
Difference to PSLC				0.35	-4.69	0.98	-3.18	2.86	4.13	8.25	-5.90	-0.04	3.49	-3.59	-5.71	0.48	-6.24	-10.48	-2.60	-6.97	-8.12	-13.38
p-value: difference to PSLC compared to LISS					*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.74*	*0.00*	*0.00*	*0.00*	*0.32*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Employment: Full																					
Point Estimate	66.50	69.70	b	71.07	61.42	70.92	82.00	61.13	65.27	58.81	62.63	65.07	65.75	55.75	55.67	59.57	64.71	53.50	58.15	56.10	60.90	84.28
Difference to LFS				1.37	-5.08	4.42	15.50	-5.37	-1.23	-7.69	-3.87	-1.43	-0.75	-10.75	-10.83	-6.93	-1.79	-13.00	-8.35	-10.40	-5.60	17.78
p-value: difference to LFS compared to LISS					*0.00*	*0.00*	*0.00*	*0.00*	*0.63*	*0.00*	*0.00*	*0.43*	*0.72*	*0.00*	*0.00*	*0.00*	*0.17*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Health: Good																					
Point Estimate	56.73	57.84	a	n/a	53.63	62.69	62.04	52.99	57.77	60.37	56.20	58.93	54.70	51.25	51.75	52.69	56.81	57.36	55.43	51.69	54.79	62.19
Difference to PSLC					-3.10	5.96	5.31	-3.74	1.04	3.64	-0.53	2.20	-2.03	-5.48	-4.98	-4.04	0.08	0.63	-1.30	-5.04	-1.94	5.46
Life satisfaction: Satisfied																					
Point Estimate	45.15	44.51	a	n/a	39.67	42.52	42.66	42.19	41.50	40.56	40.51	40.53	46.41	41.25	40.62	42.83	40.56	43.67	41.85	41.04	43.09	44.87
Difference to PSLC					-5.48	-2.63	-2.49	-2.96	-3.65	-4.59	-4.64	-4.62	1.26	-3.90	-4.53	-2.32	-4.59	-1.48	-3.30	-4.11	-2.06	-0.28
Summary employment/education																					
Average Difference				**0.86**	**4.88**	**2.70**	**9.34**	**4.12**	**2.68**	**7.97**	**4.88**	**0.73**	**2.12**	**7.17**	**8.27**	**3.71**	**4.01**	**11.74**	**5.47**	**8.68**	**6.86**	**15.58**
p-value: difference compared to LISS					***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.64***	***0.08***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
Fraction of significant statistical tests					**100%**	**100%**	**100%**	**100%**	**50%**	**100%**	**100%**	**0%**	**50%**	**100%**	**100%**	**50%**	**50%**	**100%**	**100%**	**100%**	**100%**	**100%**
Rank				**2**	**10**	**5**	**17**	**8**	**4**	**14**	**9**	**1**	**3**	**13**	**15**	**6**	**7**	**18**	**11**	**16**	**12**	**19**
Summary all non-register variables																					
Average Difference					**4.59**	**3.50**	**6.62**	**3.73**	**2.51**	**6.04**	**3.73**	**2.07**	**1.88**	**5.93**	**6.51**	**3.44**	**3.18**	**6.40**	**3.89**	**6.63**	**4.43**	**9.22**
Rank					**11**	**6**	**16**	**7**	**3**	**13**	**8**	**2**	**1**	**12**	**15**	**5**	**4**	**14**	**9**	**17**	**10**	**18**
N	**LFS: 10589**	**10632**		**5172**	**769**	**621**	**536**	**603**	**548**	**654**	**675**	**385**	**184**	**403**	**488**	**562**	**695**	**408**	**191**	**388**	**388**	**604**
	PSLC: 9607	**9499**																				

a: benchmarks comes from PSLC survey; b: benchmark comes from LFS survey

Table 5: Accuracy of Non-Register Variables, GREG estimator

	Benchmarks		Probability Panel	Non-Probability Panels																
Data Source			**LISS**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Year	**2006**	**2008**	**2008**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**	**2006**
Education: Intermediate																					
Point Estimate	36.84	37.31 [a]	38.32	34.62	39.33	34.52	42.32	40.97	47.55	33.92	37.48	40.57	33.88	30.70	37.32	31.05	25.44	33.20	30.52	29.11	21.83
Difference to PSLC			1.01	-2.22	2.49	-2.32	5.48	4.13	10.71	-2.92	0.64	3.73	-2.96	-6.14	0.48	-5.79	-11.40	-3.64	-6.32	-7.73	-15.01
p-value: difference to PSLC compared to LISS				*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.74*	*0.00*	*0.00*	*0.00*	*0.92*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Employment: Full																					
Point Estimate	66.50	69.70 [b]	70.67	58.79	73.92	82.38	58.49	62.91	54.75	62.41	63.26	65.98	53.70	55.66	59.57	63.56	56.33	57.78	57.68	59.97	80.39
Difference to LFS			0.98	-7.71	7.42	15.88	-8.01	-3.59	-11.75	-4.09	-3.24	-0.52	-12.80	-10.84	-6.93	-2.94	-10.17	-8.72	-8.82	-6.53	13.89
p-value: difference to LFS compared to LISS				*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.84*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Health: Good																					
Point Estimate	56.73	57.84 [a]	n/a	50.84	62.18	58.61	52.78	58.53	57.83	57.33	60.19	52.93	51.52	51.18	48.39	56.21	59.17	55.46	52.77	53.37	61.92
Difference to PSLC				-5.89	5.45	1.88	-3.95	1.80	1.10	0.60	3.46	-3.80	-5.21	-5.55	-8.34	-0.52	2.44	-1.27	-3.96	-3.36	5.19
Life satisfaction: Satisfied																					
Point Estimate	45.15	44.51 [a]	n/a	40.32	42.22	44.56	42.36	41.74	43.00	42.31	40.36	44.76	42.43	41.09	41.92	40.28	47.06	39.45	40.01	41.38	44.83
Difference to PSLC				-4.83	-2.93	-0.59	-2.79	-3.41	-2.15	-2.84	-4.79	-0.39	-2.72	-4.06	-3.23	-4.87	1.91	-5.70	-5.14	-3.77	-0.32
Summary employment/education																					
Average Difference			**0.99**	**4.96**	**4.96**	**9.10**	**6.75**	**3.86**	**11.23**	**3.50**	**1.94**	**2.13**	**7.88**	**8.49**	**3.71**	**4.36**	**10.78**	**6.18**	**7.57**	**7.13**	**14.45**
Difference compared to LISS																					
p-value				*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.04*	*0.11*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Fraction of significant statistical tests				100%	100%	100%	100%	100%	100%	100%	50%	50%	100%	100%	50%	100%	100%	100%	100%	100%	100%
Rank			**1**	**9**	**8**	**16**	**11**	**6**	**18**	**4**	**2**	**3**	**14**	**15**	**5**	**7**	**17**	**10**	**13**	**12**	**19**
Summary all non-register variables																					
Average Difference				**5.16**	**4.57**	**5.17**	**5.06**	**3.23**	**6.43**	**2.61**	**3.03**	**2.11**	**5.92**	**6.65**	**4.75**	**3.53**	**6.48**	**4.83**	**6.06**	**5.35**	**8.60**
Rank				**10**	**6**	**11**	**9**	**4**	**15**	**2**	**3**	**1**	**13**	**17**	**7**	**5**	**16**	**8**	**14**	**12**	**18**
N	**LFS: 10589**	**10632**	**5172**	**769**	**621**	**536**	**603**	**548**	**654**	**675**	**385**	**184**	**403**	**488**	**562**	**695**	**408**	**191**	**388**	**388**	**604**
	PSLC: 9607	**9499**																			

a: benchmarks comes from PSLC survey; b: benchmark comes from LFS survey

Table 6: Unstandardized Regression Coefficients

Independent Variables:	Health		Gender		Age	
	Nonweighted	Weighted	Nonweighted	Weighted	Nonweighted	Weighted
Constant	1.80 ***	1.80 ***	2.66 ***	2.68 ***	2.39 ***	2.42 ***
Main effect *variable*	.43 ***	.44 ***	.01	.01	.01 ***	.01 ***
Panel_1_dummy	.00	-.04	.41 ***	.35 ***	.36 ***	.31 ***
Panel_2_dummy	.05	.14	.20 ***	.19 **	.19 ***	.15 ***
Panel_3_dummy	.29 **	.21 *	.17 **	.12 **	.19 ***	.15 **
Panel_4_dummy	.05	.08	.30 ***	.29 ***	.22 ***	.21 ***
Panel_5_dummy	-.18	-.16	.06	.00	.12 **	.08 *
Panel_6_dummy	.04	.16	.27 ***	.22 ***	.32 ***	.29 ***
Panel_7_dummy	-.01	.04	.22 ***	.21 ***	.18 ***	.15 ***
Panel_8_dummy	.10	.08	.31 ***	.28 ***	.33 ***	.30 ***
Panel_9_dummy	-.06	-.06	.39 ***	.37 ***	.31 ***	.29 ***
Panel_10_dummy	-.07	.14	.43 ***	.41 ***	.37 ***	.38 ***
Panel_11_dummy	.21 *	.13	.53 ***	.48 ***	.44 ***	.42 ***
Panel_12_dummy	-.09	-.18 *	.19 **	.19 **	.31 ***	.22 ***
Panel_13_dummy	.23 **	.22 **	.15 **	.10 *	.21 ***	.18 ***
Panel_14_dummy	.11	.23	.45 ***	.39 ***	.40 ***	.33 ***
Panel_15_dummy	-.08	-.10	.36 ***	.34 ***	.42 ***	.37 ***
Panel_16_dummy	.07	.06	.43 ***	.32 ***	.53 ***	.54 ***
Panel_17_dummy	.08	-.09	.23 **	.12 *	.20 ***	.16 **
Panel_18_dummy	-.13	-.18	.13 **	.14 **	.11 **	.09 *
variable *Panel_1	.11 **	.10 **	-.13 *	-.13 *	-.17	-.19 *
variable *Panel_2	.04	-.02	-.13	-.19 **	-.31 **	-.40 **
variable *Panel_3	-.08	-.05	.00	.09	.12	.13
variable *Panel_4	.04	.03	-.14 *	-.12	.02	.15
variable *Panel_5	.11 **	.09 *	.04	.11	-.28 *	-.19
variable *Panel_6	.07	.00	.04	.10	-.25 **	-.16
variable *Panel_7	.06	.03	-.11	-.14 *	-.07	-.12
variable *Panel_8	.07	.07	.01	.04	-.08	-.03
variable *Panel_9	.12	.12	-.17	-.17	-.04	-.05
variable *Panel_10	.13 **	.04	-.16	-.16	-.23	-.37 **
variable *Panel_11	.03	.05	-.21 **	-.18 **	-.34 *	-.27 **
variable *Panel_12	.13 **	.15 ***	.11	.07	-.14	.02
variable *Panel_13	-.05	-.06	.05	.06	-.34 **	-.37 ***
variable *Panel_14	.09	.02	-.08	-.09	.07	.11
variable *Panel_15	.16 **	.15 *	.10	.06	-.10	-.05
variable *Panel_16	.14 **	.14 **	.25 **	.41 ***	.02	-.16
variable *Panel_17	.01	.06	-.11	-.04	-.36 *	-.44 **
variable *Panel_18	.12 *	.14 **	-.05	-.09	.15	.01

* *p<.10; **p<.05; ***p<.001*

Table 7: Spearman Rank Correlations of Panel Management Techniques with Deviations from Benchmarks and Probability Face-to-Face Samples

	Register variables	Nonregister variables			
		Nonweighted		Weighted	
Panel management	average Δ benchmark	average Δ benchmark		average Δ benchmark	
Number of active Members	.11	.00		.02	
Age of panel	-.14	-.18		-.16	
Panel refreshment	-.32	-.22		-.22	
Number of invitations per months	-.23	.10		-.04	
RECRUITMENT INTO PANEL THROUGH:					
Traditional modes	.20	-.03		-.04	
Links on websites	.12	-.54	*	-.65	**
Telephone	-.01	-.47		-.57	**
Bought addresses	.25	.26		.27	
Snowball method	-.59	-.01		.00	
Existing mail panels	-.31	-.04		-.05	
INCENTIVE SYSTEM					
Point system	.34	-.20		-.25	
Prizes through lottery	-.42	.12		.12	
Monetary incentives	-.39	-.08		.03	
Lottery tickets for lottery	.39	-.48		-.48	
Use of reward for registering in panel	.39	-.39		-.52	*

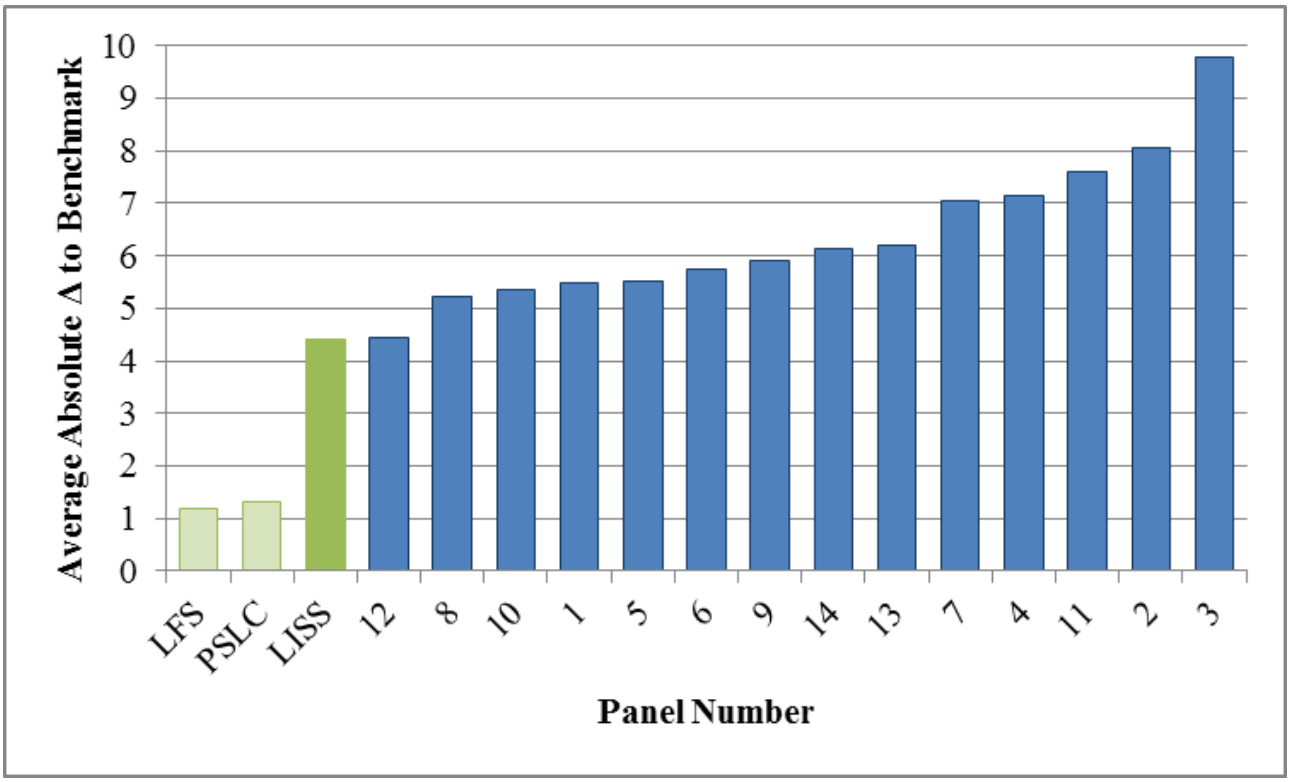
** significant at p<.10; ** significant at p<.05; ***significant at p<.001*

Table 7: Overview of studies published in JM and JMR which conduct surveys

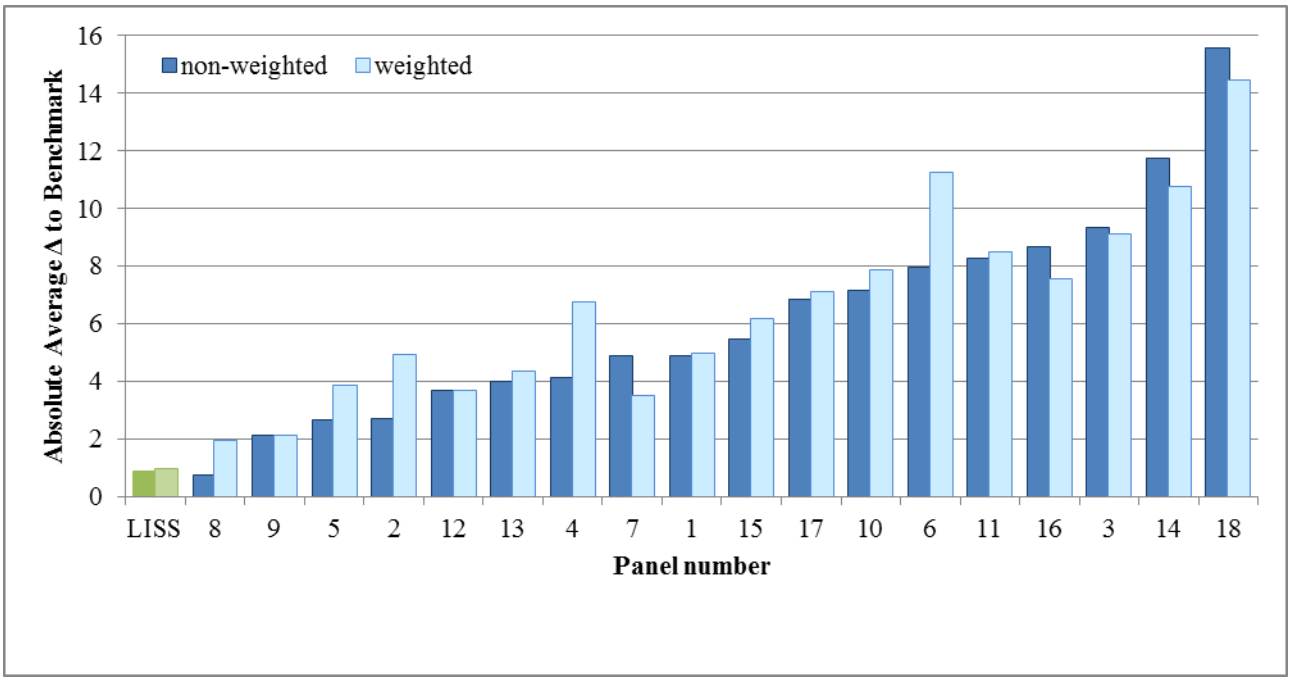
	JM: 2009-2013 % of studies that report this information	JMR: 2009-2013 % of studies that report this information
Sampling frame	**96%**	**99%**
Students	10%	19%
Company	56%	33%
(Online) panel	10%	7%
Mall intercept	5%	7%
Experts	1%	12%
Database	14%	14%
Other	0%	7%
Sampling mode	**86%**	**66%**
Mail	31%	26%
Telephone	4%	5%
Internet	27%	21%
Face-to-face	15%	14%
Sampling method	**36%**	**39%**
Simple random sampling	23%	35%
Stratified sampling	1%	0%
Convenience sampling	7%	2%
Quota sampling	0%	2%
Snowball sampling	5%	0%
Response rate	**77%**	**49%**
Reliability	87%	53%
Validity	**77%**	**56%**
- face	14%	9%
- content	10%	0%
- predictive	7%	9%
- concurrent	0%	0%
- convergent	43%	30%
- discriminant	70%	44%
Errors	**69%**	**44%**
- nonresponse	55%	33%
- sampling	1%	2%
- coverage	0%	0%
- measurement	37%	33%
Response quality	**18%**	**16%**
- response style	1%	2%
- response effort	0%	2%
- accuracy	11%	9%
- inconsistency	0%	0%

Figure 1: Average Difference to Benchmark

a) Register variables

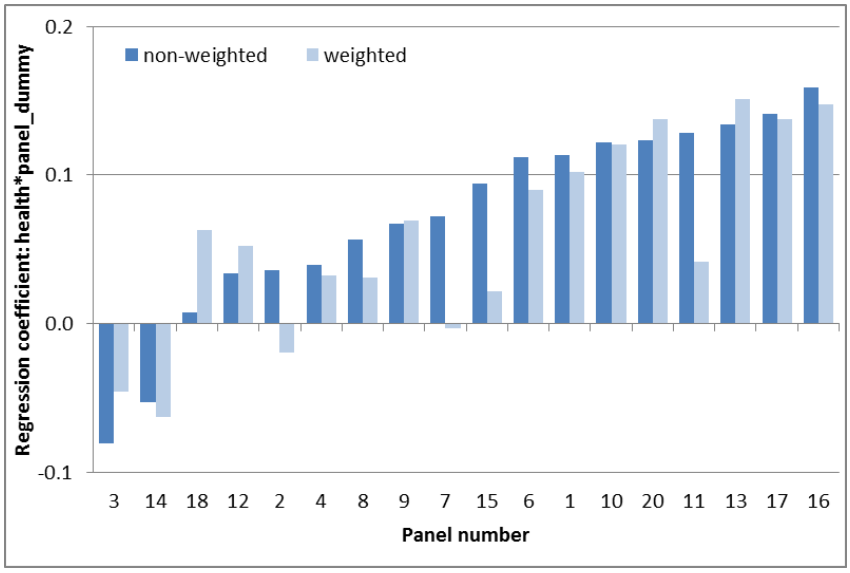


b) Non-register variables

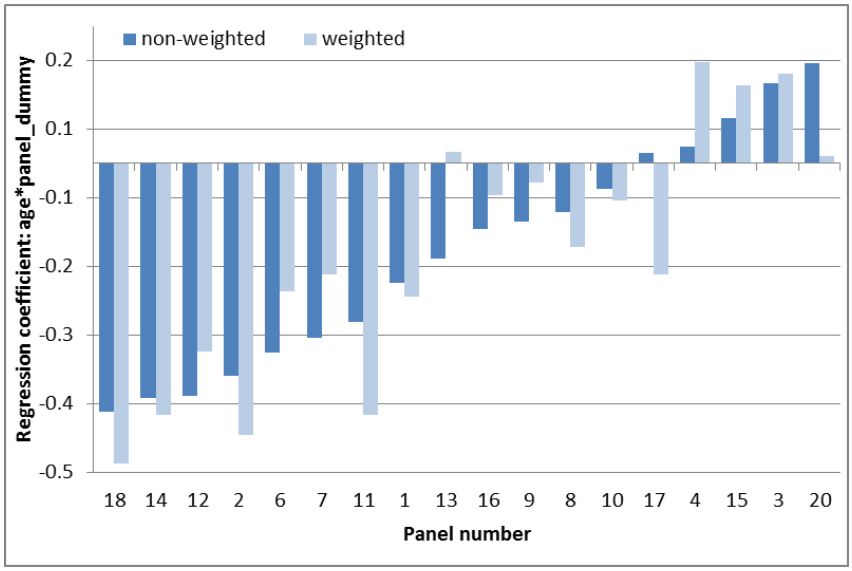


Non-register variables: Results are presented for Employment and Education.

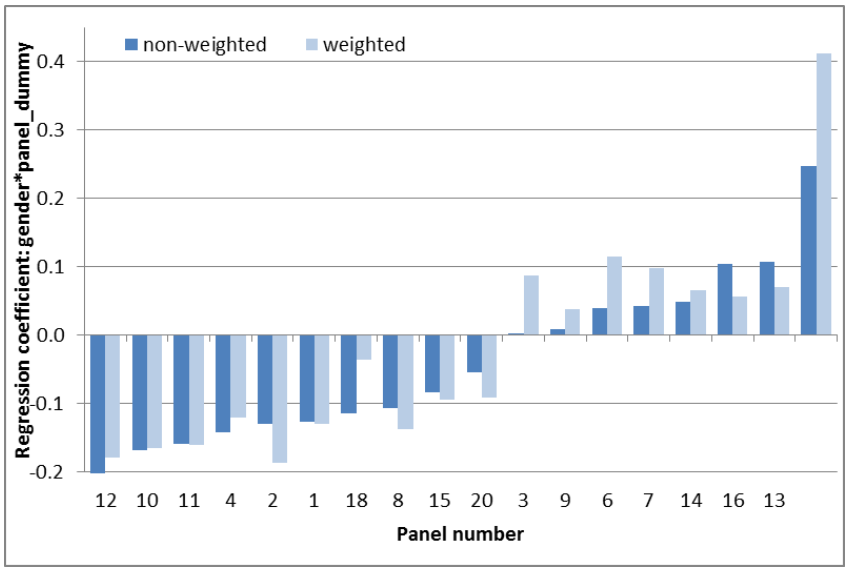
Figure 2: Non-weighted and weighted Unstandardized Regression Coefficients



Life satisfaction on health



Life satisfaction on age



Life satisfaction on gender

Variable	Variable	MBA and Probability Face-to-Face Surveys[a]		NOPVO		LISS	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
Geslacht	**Gender**	**Geslacht: ○ M ○ V**	**Gender: ○ M ○ F**	**Geslacht: ○ Man ○ Vrouw**	**Gender: ○ Male ○ Female**	**Geslacht: ○ Man ○ Vrouw**	**Gender: ○ Male ○ Female**
		Man	Male	Man	Male	Man	Male
		Vrouw	Female	Vrouw	Female	Vrouw	Female
Nationaliteit	**Nationality**	***afgeleid uit geboorteaangifte (verplicht voor kinderen geboren in NL) of geldig legitimatiebewijs (voor immigranten)***	***derived from birth register (obligatory for children born in the Netherlands) or valid ID (for immigrants)***	**Welke nationaliteit heeft u?**	**What is your nationality?**	**Wat is uw nationaliteit?**	**What is your nationality?**
		Nederlandse	Dutch	Nederlandse	Dutch	Nederlandse	Dutch
		Niet Nederlandse	Non-Dutch	Niet Nederlandse	Non-Dutch	Niet Nederlandse (Turkije, Marokko, Nederlandse Antillen, Suriname, Indonesië, ander niet-westers land, ander westers land)	Non-Dutch (Turkey, Morocco, Netherlands Antilles, Suriname, Indonesia, other non-western country, another western country)
Geboorteland zelf/ moeder /vader	**Own/mother's /father's country of origin**	***afgeleid uit geboorteaangifte (verplicht voor kinderen geboren in NL) of geldig legitimatiebewijs (voor immigranten)***	***derived from birth register (obligatory for children born in the Netherlands) or valid ID (for immigrants)***	**In welk land bent u geboren? En uw vader? En uw moeder?**	**In which country were you born? And your father? And your mother?**	**In welk land bent u geboren? In welk land werd uw vader geboren? In welk land werd uw moeder geboren?**	**In which country were you born? In which country was your father born? In which country was your mother born?**
		Nederland	Netherlands	Nederland	Netherlands	Nederland	Netherlands
		Anders	Other	Anders (Suriname, Nederlandse Antillen/Aruba, Indonesië, Turkije, Marokko, Anders)	Other (Suriname, the Netherlands Antilles / Aruba, Indonesia, Turkey, Morocco, Other)	Anders (Turkije, Marokko, Nederlandse Antillen, Suriname, Indonesië, ander niet-westers land, ander westers land)	Other (Turkey, Morocco, Netherlands Antilles, Suriname, Indonesia, other non-western country, another western country)
Aantal personen in huishoud	**Number of people in household**	**aantal personen woonachtig op het adres [+aangever moet zichzelf opvoeren en naam, geslacht, een relatie voor alle inwonenden opgeven]**	**number of persons residing at the address [+ declarant must list him/herself and specify name, gender, and relationship to all residents]**	**Uit hoeveel personen bestaat uw huishouden, u zelf meegerekend?**	**How many people live in your household, including you yourself?**	**Uit hoeveel personen bestaat uw huishouden, inclusief uzelf?**	**How many people live in your household, including you yourself?**
		1	1	1	1	1	1
		2	2	2	2	2	2
		3	3	3	3	3	3
		4	4	4	4	4	4
		5	5	5	5	5	5
		6 of meer	6 or more	6 of meer	6 or more	6 of meer	6 or more
Leeftijd	**Age**	**Geboortejaar**	**Year of birth**	**Wat is u geboortejaar?**	**What is your year of birth?**	**Geboortejaar**	**Year of birth**
		17-24	17-24	17-24	17-24	17-24	17-24
		25-34	25-34	25-34	25-34	25-34	25-34
		35-44	35-44	35-44	35-44	35-44	35-44
		45-54	45-54	45-54	45-54	45-54	45-54
		55-66	55-66	55-66	55-66	55-66	55-66

Variable	Variable	MBA and Probability Face-to-Face Surveys[a]		NOPVO		LISS	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
Regio	**Region**	***afgeleid van postcode***	***derived from postal code***	***afgeleid van postcode***	***derived from postal code***	***afgeleid van postcode***	***derived from postal code***
		3 grote steden	3 large cities	3 grote steden	3 large cities	3 grote steden	3 large cities
		Oost	East	Oost	East	Oost	East
		Noord	North	Noord	North	Noord	North
		Zuid	South	Zuid	South	Zuid	South
		West	West	West	West	West	West
Provincie	**Province**	***afgeleid van postcode***	***derived from postal code***	***afgeleid van postcode***	***derived from postal code***	***afgeleid van postcode***	***derived from postal code***
		Groningen	Groningen	Groningen	Groningen	Groningen	Groningen
		Friesland	Friesland	Friesland	Friesland	Friesland	Friesland
		Drenthe	Drenthe	Drenthe	Drenthe	Drenthe	Drenthe
		Overijssel	Overijssel	Overijssel	Overijssel	Overijssel	Overijssel
		Flevoland	Flevoland	Flevoland	Flevoland	Flevoland	Flevoland
		Gelderland	Gelderland	Gelderland	Gelderland	Gelderland	Gelderland
		Utrecht	Utrecht	Utrecht	Utrecht	Utrecht	Utrecht
		Noord-Holland	North Holland	Noord Holland	North Holland	Noord Holland	North Holland
		Zuid-Holland	South Holland	Zuid Holland	South Holland	Zuid Holland	South Holland
		Zeeland	Zeeland	Zeeland	Zeeland	Zeeland	Zeeland
		Noord-Brabant	North Brabant	Noord Brabant	North Brabant	Noord Brabant	North Brabant
		Limburg	Limburg	Limburg	Limburg	Limburg	Limburg

[a] All people residing in the Netherlands are obliged to register at their local municipality, where information on socio-demographic variables is verified based on official documents (e.g. valid ID, passport, marriage certificate). People are also obliged to register changes (e.g. birth of children, relocation, divorce). The information from all the municipalities in the Netherlands is sent to Statistics Netherlands on a daily basis.

Appendix B: Wordings of Questions Measuring Non-Register Variables

Variables		Probability Face-to-Face Surveys			NOPVO		LISS	
Dutch	**English**		**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
Opleiding	**Education**	**LFS**	**Heft u na de lagere school of basisschool een opleiding of cursus gevolgd waarmee u 2 jaar of langer bezig bent gewest? Welke opleiding of cursus was dat? (==> waarvan de hoogsde opleiding werd afgeleid)**	**What is the highest level of education that you have attended?**	**Wat is de hoogste opleiding die u gevolgd hebt?**	**What is the highest level of education that you have attended?**	**Wat is de hoogste opleiding die u gevolgd hebt?**	**What is the highest level of education that you have attended?**
			basisonderwijs	primary school	geen onderwijs/	primary school	geen onderwijs/ basisschool	primary school
			VMBO, mbo1, AVO	primary education	LBO\VMBO (VBO)	primary education	VMBO	primary education
			HAVO, VWO, MBO	secondary education	HAVO, VWO, MBO	secondary education	HAVO, VWO, MBO	secondary education
			HBO, WO	higher education	HBO, WO	higher education	HBO, WO	higher education
Werk	**Employment**	**LFS**	**De volgende vragen gaan over u huidige situatie op de arbeidsmarkt. Heeft u op dit moment betaald werk (1)? Als nee: Zou u op dit moment betaald werk willen hebben (2)? Als nee: Wat is de voornamste reden**	**The following questions are about your current situation on the labor market. Do you currently do paid work (1)? If no: Would like to have paid work at this moment (2)? If not: What is the grandest reason why you do**	**Tot welke groep rekent u zichzelf?**	**To which group do you count yourself?**	**Kunt u aangeven welke van de volgende omschrijvingen op u van toepassing zijn?**	**Can you indicate which of the following descriptions applies to you?**
			Werkt tenminste 12 u/wk (Werkzame beroepsbevolking)	Works at least 12 hrs / wk (Employed labor force)	Doet betaald werk voor meer dan 15 uur per week	Doing paid work for more than 15 hours per week	Verricht betaald werk in loondienst, Werkt of is meewerkend in gezins- of familiebedrijf, Is vrije beroepsbeoefenaar,	Paid employment, works or assists in family business, autonomous professional, freelancer or self-employed
			Wil werken	Job seeker	Werkloos na verlies werkkring, zoekt voor het	Job seeker following job loss, first-time job seeker	Zoekt werk na verlies werkkring, zoekt voor het	Job seeker following job loss, first-time job seeker
			Wil/kan niet werken vanwege zorg voor gezin of	Does not want to/ can not work due to care for family	Doet eigen huishouden	Takes care of the housekeeping	Verzorgt de huishouding	Takes care of the housekeeping
			Wil/kan niet werken vanwege ziekte/ arbeidsongeschiktheid of slechte gezondheid	Does not want to / can not work because of illness / disability or poor health	Arbeidsongeschikt, WAO, AAW	Work disability	Is (gedeeltelijk) arbeidsongeschikt	Has (partial) work disability

			Wil/kan niet werken vanwege vut/pensioen of hoge leeftijd	Does not want to / can not work because of early retirement / pension or old	Gepensioneerd, in de VUT	Pensioner, early retirement	Is met pensioen (vervroegd, AOW of VUT)	Is pensioner ([voluntary] early retirement, old age pension scheme)
			Wil/kan niet werken vanwege opleiding/studie	Does not want to / can not work because of education /	Scholier, student	Schoolchild, student	Gaat naar school of studeert	Attends school or is studying
			Wil/kan niet werken maar wil werk <12 u/wk vanwege andere redenen	Does not want to / can not work but wants work for other reasons	Anders	Other	ANDERS: Verricht onbetaald werk met behoud van uitkering, Verricht vrijwilligerswerk, Doet iets anders, Is te jong, heeft nog geen bezigheden, Vrijgesteld van werkzoeken na verlies van werkkring	OTHER: Performs unpaid work while retaining unemployment benefits, performs voluntary work, does something else, is too young to have an occupation, exempted from job seeking following job loss
Gezondheid	**Health**	**PSLC**	**Hoe is over het algemeen uw gezondheid?**	**How is your health, generally speaking?**	**Hoe is over het algemeen uw gezondheid?**	**How would you describe your health, generally speaking?**	***Hoe zou u over het algemeen uw gezondheid noemen?***	***How would you describe your health, generally speaking?***
			Zeer goed	Very good	Zeer goed	Very good	*Uitstekend*	*Excellent*
			Goed	Good	Goed	Good	*Zeed goed*	*Very good*
			Gaat wel	Decent	Gaat wel	Decent	*Goed*	*Good*
			Slecht	Bad	Slecht	Bad	*Matig*	*Moderate*
			Zeer slecht	Very bad	Zeer slecht	Very bad	*Slecht*	*Bad*
Tevredenheid leven	**Life satisfaction**	**PSLC**	**In welke mate bent u tevreden met het leven dat u op dit moment leidt?**	**To what extent are you satisfied with the life you currently lead?**	**In welke mate bent u tevreden met het leven dat u op dit moment leidt?**	**How satisfied are you with your life right now?**	***In welke mate bent u tevreden met het leven dat u op dit moment leidt?***	***How satisfied are you with the life you lead at the moment?***
			Buitengewoon tevreden	Extremely satisfied	Buitengewoon tevreden	Extremely satisfied	*0-10, "helemaal ontevreden - helemaal tevreden"*	*0 - 10, "not at all satisfied - completely satisfied"*
			Zeer tevreden	Very satisfied	Zeer tevreden	Very satisfied		
			Tevreden	Satisfied	Tevreden	Satisfied	*Because of the different scales, we did not include health and life satisfaction for LISS in our analyses*	
			Tamelijk tevreden	Failry satisfied	Tamelijk tevreden	Failry satisfied		
			Niet zo tevreden	Not so satisfied	Niet zo tevreden	Not so satisfied		

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2015–2016	2015 to 2016 inclusive
2015/2016	Average for 2015 to 2016 inclusive
2015/'16	Crop year, financial year, school year, etc., beginning in 2015 and ending in 2016
2013/'14–2015/'16	Crop year, financial year, etc., 2013/'14 to 2015/'16 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Studio BCO, Den Haag

Design
Edenspiekermann

Information
Telephone +31 88 570 7070
Via contact form: www.cbs.nl/information