Proceedings of Statistics Canada Symposium 2014
Beyond traditional survey taking: adapting to a changing world

Using Web Panels for Official Statistics

Jelke Bethlehem[1]

Abstract

New developments in computer technology, but also new challenges in society like increasing nonresponse rates and decreasing budgets may lead to changes in survey methodology for official statistics. Nowadays, web panels have become very popular in the world of market research. This raises the question whether such panels can also be used for official statistics. Can they produce high quality statistics about the general population? This paper attempts to answer this question by exploring methodological aspects like under-coverage, sample selection, and nonresponse. Statistics Netherlands carried out a test with a web panel. Some results are described.

Key Words: Web panel, representativity, under-coverage, nonresponse, sampling.

1. Introduction

1.1 Data collection for official statistics

It is the task of National Statistical Institutes (NSI's) to produce accurate statistics. Traditionally, they conducted face-to-face or telephone surveys to collect data for these statistics. This is an expensive way of data collection, but experience has shown that it is the price that has to be paid for obtaining high quality data. Nowadays, budget constraints cause NSI's in many countries to look for less expensive ways of data collection while maintaining a high level of data quality.

At first sight, a web panel may seem a promising alternative. Online data collection has become increasingly popular, particularly in the world of market research. This is not surprising, as it is a simple, fast and cheap way to collect large amounts of data. Once a web panel has been put into place, it is simple to conduct a survey. No complex and expensive sample selection procedures are required. It is just a matter of sending an email to (a sample of) panel members. No interviewers are needed, and there are no mail costs for sending paper questionnaires. It suffices to put the electronic questionnaire on the internet.

Speed is another advantage of online data collection. A new survey can be launched quickly. There are examples of web surveys for which questionnaire design, data collection, analysis and publication took no more than just one day. Combining this advantage with its longitudinal nature, web panels have become a powerful tool for opinion polls. For example, in the last weeks of the campaign for the parliamentary elections of 2012 in the Netherlands, there were four different major national polls each day, and they were all based on web panels. See Bethlehem (2013).

Web panels can be used in various ways in official statistics:

- For longitudinal studies, where the same set of variables is measured for the same group of individuals at different points in the time. The focus is on measuring change.
- For cross-sectional studies, where the panel serves as a sampling frame for specific surveys that may address different topics. Samples may be selected from different groups (elderly, high-educated, etc.).
- For quick ad-hoc (unplanned) surveys that an NSI may conduct on request by third parties.

[1]Jelke Bethlehem, Leiden University, Faculty of Social and Behavioural Sciences, Wassenaarseweg 52, 2333 AK Leiden, the Netherlands (j.g.bethlehem@fsw.leidenuniv.nl)

This paper concentrates on the cross-sectional use of web panels for the general population. To allow for proper statistical inference about a population, recruitment and sampling must be based on probability sampling. Examples of such panels are the LISS panel in the Netherlands (Scherpenzeel, 2008) and the KnowledgePanel in the US (Knowledge Networks, 2012) and the ELIPSS Panel in France (Lynn, 2013).

The paper explores the possible use of web panels for official statistics. Can web panels produce statistics that are as accurate as those produced by CAPI surveys? An attempt is made to answer this question by addressing a number of issues, such as under-coverage, recruitment, and nonresponse.

2. Web panels

2.1 Under-coverage

A survey taken from a web panel may suffer from under-coverage because the target population of a survey is usually much wider than just people with internet. For example, according to Eurostat, the statistical office of the European Union, 79% of the households in the EU had access to internet in 2013. There were large differences between countries. The countries with the highest internet coverage were Iceland (96%), the Netherlands (95%), Norway (94%), and Luxemburg (94%). Internet access was lowest in Turkey (49%), Bulgaria (54%), Greece (56%), and Romania (58%). Even more problematic is that internet access is unevenly distributed over the population. A typical pattern found in many countries is that elderly, low-educated and ethnic minorities are severely under-represented among those having internet.

Bethlehem & Biffignandi (2012, chapter 8) show that the bias due to under-coverage of the sample mean as an estimator of the population mean of a target variable is determined by two factors:

- The percentage of people with internet access. The higher internet coverage, the smaller the bias.
- The contrast between the people with and without internet. This is the difference between the population means of those with and without internet. The more the means of the target variable differ between these groups, the larger the bias will be.

The percentage of people without internet cannot be neglected in many countries. Moreover, there are substantial differences between those with and without internet. Specific groups are under-represented in the internet population. So, the conclusion is that generally a random sample from the internet population will lead to biased estimates of the parameters of the target population.

It is to be expected that internet coverage will increase over time. Hence, the percentage of people with internet will increase, and this will reduce the bias. It is unclear, however, whether the contrast will also decrease over time. It is even possible that it increases, as the group without internet may differ more and more from the group with internet. So the combined effect of more people with internet and a larger contrast need not necessarily lead to a smaller bias.

One way to solve the under-coverage problem is to provide free internet access to sample persons without internet. This approach was implemented for the KnowledgePanel in the US, and the LISS panel in the Netherlands. Leenheer & Scherpenzeel (2013) show that this improved the representativity of the LISS panel. The surveyed population was closer to the general population. Lynn (2013) mentions the French ELIPSS panel in which every member was provided with a tablet and a 3G subscription. This approach has the advantage that every panel member uses the same device for completing the questionnaires, thereby avoiding device effects.

The approach of providing panel members with internet devices also raises new questions. Is this approach still feasible if the under-coverage is substantial? And what about the people who have no experience at all with the internet? Will this lead to incomplete interviews, or measurement errors? A different solution of the under-coverage problem, is to extend the web panel to a mixed-mode panel. Sample persons without internet access will be approached in a mode different than web (mail, CATI or CAPI). Whatever approach is used to do reduce under-coverage (offering free internet access or setting up a mixed-mode panel), it will increase the costs of the panel.

2.2 Panel recruitment

Setting up a web panel that allows for valid statistical inference about a general population, requires selection of a probability sample. This is not easy, because usually there is no proper sampling frame available. Therefore, many web panels rely on some form of self-selection. Self-selection (also called opt-in) means that it is completely left to people to select themselves for the panel. Respondents are those who happen to have internet, encounter an invitation, visit the appropriate website, and decide to participate.

In case of self-selection, the researcher is not in control of the selection process. Each person has an unknown participation probability, which makes it impossible to construct unbiased estimators. Another problem is that also people from outside the target population can become panel member. Moreover, people can have multiple membership (possibly under different identities). Another problem is that people may try to manipulate the outcome of the survey. Bronzwaer (2012) gives an example. During the campaign for the parliamentary elections in the Netherlands in 2012 there was a group of 2,500 people who applied for membership of a web panel (Peil.nl) that conducted political polls. Their idea was to first behave as voters for the old people party 50PLUS, and then gradually change to the Christian-democrats. Unfortunately for them, their scheme was detected when the polling organization observed a sudden, large increase in the number of people volunteering to become a panel member. Nevertheless, this anecdote shows that manipulation of self-selection web panel is technically possible.

To show the effects of self-selection on estimators, it is assumed that each person in the internet population has an unknown probability of participating in a survey. A naive researcher assuming simple random sampling, will use the sample mean. Bethlehem & Biffignandi (2012, chapter 9) show that this estimator is biased, and this bias is determined by three factors:

- The average participation probability. The larger this average participation probability is, the smaller the bias will be.
- The variation of the participation probabilities. The more these probabilities vary, the larger the bias will be.
- The correlation between the target variable and the participation probabilities. The stronger the correlation, the larger the bias.

It should be noted that the average participation probability can be very small, and thus the bias very large. For example, there are web panels in the Netherlands with 100,000 members. The target population (all Dutch from the age of 18) consists of 12.8 million people. Hence the average participation probability is equal to 100,000 / 12,800,000 = 0.008.

The bias vanishes if all participation probabilities are equal. The self-selection mechanism is then comparable to a simple random sample. The bias also vanishes if participation does not depend on the value of the target variable.
A self-selection panel is considered out of the question for compiling accurate statistics about the general population. Indeed, a special task force of AAPOR (American Association of for Public Opinion Research) concluded that “Researchers should avoid nonprobability online panels when one of the researcher objectives is to accurately estimate population values”, see Baker et al. (2010).

Ideally, the sampling frame for a web panel is a list of e-mail addresses of all people in the target population. Such a list could exist e.g. for all students of a university, or for all employees of a large company, but unfortunately there is no such list for the general population. The way out is to use another mode of recruitment. Here are some examples:

- Select a random sample from a population register, and send the selected persons a letter containing the internet address of the survey and a unique login code. This approach is currently used by Statistics Netherlands to conduct web surveys and mixed-mode surveys.
- Select a random sample from a population register, use telephone directories to link as much as possible telephone numbers to the selected persons, call these persons and invited them to become member of the panel. If a telephone number cannot be linked, attempt to get cooperation by means of a CAPI interview. An approach like this was used for the LISS panel, see Leenheer & Scherpenzeel (2013).
- Select a random sample of telephone numbers (for example by means of random digit dialling), call each selected number and invite a random household member to become a panel member.

The first two recruitment approaches are only possible if a population register is available, like in the Scandinavian countries and the Netherlands. If this is not the case, a similar approach may be to use an address list, for example a postal address file (PAF).

Large survey organizations conducting many surveys could consider recruiting a panel from the response of one of their CAPI or CATI surveys. For example, Statistics Netherlands always asks its CAPI/CATI respondents whether they want to participate again in some future survey. Cooperative persons may also want to become a panel member. Moreover, sometimes their e-mail addresses are already available so that use of a different recruitment mode can be avoided. However, this approach also has the important disadvantage that the CAPI/CATI respondents do not constitute a random sample from the general population. This may lead to a panel that lacks representativity, and thus to biased estimates.

2.3 Nonresponse

Nonresponse as an important issue for web panels. It occurs in two phases of a web panel: (1) during the recruitment phase, and (2) in the specific surveys taken from the panel. Recruitment nonresponse may be high because participating in a web panel requires substantial commitment and effort of respondents. The response rates of web surveys based on probability sampling are often below 40%. See, for example, Cook, Heath & Thompson (2000), Kaplowitz, Hadlock & Levine (2004), and Lozar Manfreda et al. (2008). The response rate of a specific survey taken from the panel is often high as the panel members all agreed to do a survey regularly.

Nonresponse can have various causes. These causes depend on the recruitment approach used to set up the panel. If a sampling frame with e-mail addresses is available, nonresponse can be caused by non-contact (wrong e-mail address, message was intercepted by a spam filter, message was not read), by refusal (after reading the e-mail message), or not able (e.g. browser not working properly). If recruitment is by mail or telephone, the causes of nonresponse are similar to those for mail or telephone surveys.

Attrition is a specific type of nonresponse that may occur in the surveys taken from a panel. People can get tired of completing the specific survey questionnaires and therefore they may decide to stop their cooperation. Once they stop, they will never start again.

The problem of nonresponse is that it may be selective. To show what the effects of nonresponse can be, it is usually assumed that every person in the population has an unknown response probability. Then, according to Bethlehem, Cobben & Schouten (2011), the bias of the response mean (as an estimator of the population mean) is determined by three factors:

- The average response probability. The higher the average response probability, the smaller the bias.
- The variation of the response probabilities. The more these probabilities vary, the larger the bias will be.
- The correlation between target variable and response probabilities. The stronger the correlation, the larger the bias.

The bias vanishes if all response probabilities are equal. Then, the response can be seen as a simple random sample. The bias also vanishes if response probabilities does not depend on the value of the target variable.

The mode of recruitment has an impact on the recruitment response rate. Typically, interviewer-assisted surveys (CAPI, CATI) have higher response rates than self-administered surveys (mail, web). From the point of view of representativity, interviewer-assisted recruitment should therefore be preferred. However, this increases recruitment costs considerably.

It is important to keep to the probability sampling paradigm: samples must be selected by means of probability sampling, and the selection probabilities must be known. Self-selection (opt-in) was rejected as a scientifically sound sampling technique, because selection probabilities are unknown. Moreover, these probabilities may be 0 for some groups. But what about a proper probability sample that is affected by a substantial amount of nonresponse? Does this not almost resemble a self-selection survey? Bethlehem (2010) shows that the expressions for the nonresponse bias and the self-selection bias are similar, be it that the expression for the nonresponse bias contains response probabilities and that for the self-selection bias contains participation probabilities. Response probabilities are generally much higher than participation probabilities. This means the self-selection bias can be much larger than the nonresponse bias. As an example, the recruitment response rate of the probability sampling based LISS Panel was 54%, corresponding to a mean response probability of 0.54. The participation rate in one Dutch self-selection web panel was 0.008. This implies that, in the worst case, the self-selection bias can be more than 12 times as large as the nonresponse bias.

The response rate is not the only factor determining the nonresponse bias. Another important factor is the variation of the response probabilities. This is measured by their standard deviation. If all response probabilities are equal, the standard deviation is 0, and there is no bias. The more the response probabilities vary, the larger the bias. Schouten, Cobben & Bethlehem (2009) propose the R-indicator as a measure of representativity. This indicator is defined as $R = 1 - 2S_\rho$, where S_ρ is the standard deviation of the response probabilities. R is equal to 1 if all response probabilities are the same. This is the case of complete representativity. The closer the value of R is to 0, the larger the lack of representativity is.

Table 2.3.1 gives an example of nonresponse in a web panel. The data relate to the LISS Panel and are taken from Scherpenzeel & Schouten (2011). The table shows the response rates in the subsequent phases of the process. Percentages are with respect to the initial sample. Contact could be established with 91% of the sample households. In 75% of the cases households agreed to do a short recruitment interview. 54% of the households decided to participate in the panel, but only 48% became active in the panel. So, the response rate of the recruitment phase was 54%. The table also shows the effect of attrition. The response rate decreases over time. After four years, only one out of three original sample households is still active in the panel.

Table 2.3.1

Response rates and R-indicator for the LISS Panel

Phase	Response	R-indicator
Recruitment contact	91 %	0.85
Recruitment interview	75 %	0.80
Agree to participate in panel	54 %	0.71
Active in panel in 2007	48 %	0.67
Active in panel in 2008	41 %	0.70
Active in panel in 2009	36 %	0.75
Active in panel in 2010	33 %	0.78

The R-indictor is still high in the contact phase (0.85), but it decreases in the course of the recruitment process. The R-indicator is only 0.67 when the panel starts in 2007, indicating the panel is not so representative. Surprisingly, the R-indicator increases again over the years. Apparently, attrition causes the composition of the panel to become more balanced.

2.4 Bias correction

Representativity of a web panel can be affected by at least three phenomena: under-coverage, self-selection, and nonresponse. Lack of representativity may lead to biased estimates. Weighting adjustment is a family of correction techniques that attempt to reduce the bias by using auxiliary variables. Auxiliary variables are defined here as variables that are measured in a survey, and for which their population distribution is available.

By comparing the response distribution of an auxiliary variable with its population distribution, it becomes clear whether or not the response is representative for the population (with respect to this variable). If both distributions differ considerably, one must conclude that the response is selective. To correct for this, adjustment weights are computed for the respondents. Respondents in an under-represented group get a weight larger than 1, and those in an over-represented group get a weight smaller than 1. Estimates of population characteristics are then computed by using weighted data instead of unweighted data.

Post-stratification is a simple and often used weighting technique. More advanced weighting techniques are generalized regression estimation and raking ratio estimation. There are also weighting techniques using estimated response probabilities. This is called propensity score weighting. A more in-depth treatment of weighting techniques can be found in, for example, Bethlehem & Biffignandi (2012) and Särndal & Lundström (2005). All these weighting techniques are only capable of reducing the selection bias if the auxiliary variables used satisfy two conditions:

- The auxiliary variables must have a strong correlation with the response/participation probabilities. They must be able to explain the selection mechanism;
- The auxiliary variables must have a strong correlation with the target variables of the survey/panel. They must have sufficient power to explain the values of the target variables.

The bias will vanish completely if all auxiliary variables are used that are required to completely explain the selection mechanism and the target variables. If only a subset is used, bias will be reduced but not removed.

Selection problems occur in the two phases of the web panel. During recruitment there can be under-coverage, self-selection and nonresponse. And there can be nonresponse in the surveys taken from the web panel. Ideally, bias correction should also take place in two steps. Recruitment selectivity is a different phenomenon than survey nonresponse. Therefore it may require a different model containing different auxiliary variables. Moreover, there are a lot more auxiliary variables available to correct for the survey nonresponse bias. For many web panels, new members conduct a profile survey. This is an initial questionnaire asking basic demographic questions. All these variables can be used to weight the survey response data. There are often less auxiliary variables available for weighting adjustment in the recruitment phase. To summarize, weighting adjustment in a web panel is a two-step process:

(1) Compute weights for all panel members in such a way that the panel becomes representative with respect to the target population.
(2) For each survey, compute weights in such a way that the survey becomes representative with respect to the panel.

The final weights are obtained by multiplying the recruitment weights by the survey weights.

3. A web panel pilot

3.1 Recruitment

Statistics Netherlands carried out a pilot with a web panel in 2012. Main objective to get some first experiences with building web panels. There was little time available and resources were limited. For this reason it was decided to recruit panel members from the respondents of an existing survey: OViN (Onderzoek Verplaatsingen in Nederland). It is a mobility survey. The sample for this survey was selected at random from the population register. At the end of the questionnaire, all respondents were asked whether they would be willing to participate in other Statistics Netherlands surveys. Those with a positive response (and having access to internet) were invited by mail to become a member of the web panel for a period of one year, and to complete one questionnaire each month.

Statistics Netherlands could link the OViN sample to several registrations. Hence, there was a large set of auxiliary variables. The values of these variables became available for both panel members and people not in the panel. Therefore, these auxiliary variables allowed for analysis of the representativity of the panel.

The sample for the web panel was selected from the OViN respondents. Therefore the target population of the web panel were the OViN respondents, and not the general population. This was taken into account when investigating the representativity of the web panel.

The recruitment process for the web panel is summarized in table 3.1.1. The original sample for OViN consisted of 12,406 persons selected at random from the population register. The OViN response rate was 57.5%, which corresponds to 6,928 persons. Of those respondents, 4,251 (35.3% of the original sample) agreed to participate in the panel. Only 1,231 willing people really registered for the panel, and only 1,134 willing people really completed the questionnaire of the first survey taken from the panel. So in the end, only 9.4% of the initial OViN sample became active in the panel. This is a low score, which raised concerns about representativity.

Table 3.1.1 contains the R-indicator for several steps in the recruitment process. The OViN response lacked some representativity as the R-indicator was 0.78. Surprisingly, the R-indicator increased to 0.88 in the subsequent steps. The remaining group of people became more representative as more people dropped off. So the final composition of the panel was more balanced than the initial group of OViN respondents.

Table 3.1.1

The recruitment process

Phase	Size	Response rate	R-indicator
OViN sample	12,406		
OViN response	6,928	57.5%	0.78
Willing to enter panel	4,251	35.3%	0.84
Selected for panel	4,227	35.1%	0.84
Registered in panel	1,231	10.2%	0.88
Participated in panel	1,134	9.4%	

The usability of the web panel was explored by analysing two target variables: *level of education*, and *main activity*. All values of these two variables were available because the OViN sample could be linked to administrative sources. Table 3.1.2 compares the distribution in the panel of level of education with the corresponding distribution in the OViN sample (the target population). The percentage of low educated in the panel is only 2.6%, whereas it should be much larger (5.5%). High educated people seem to be over-represented in the panel. The percentage of high educated in the panel is 45.5%, whereas it should be only 33.6%.

Table 3.1.2

Estimates (unweighted and weighted) for level of education (%)

Level of education	Panel	Weighted	OViN
Primary	2.6	4.3	5.5
Lower secondary	15.2	16.5	21.0
Higher secondary	34.4	35.8	37.6
Bachelor/master	45.5	40.6	33.6

To explore whether weighting could improve estimates, a weighting model was constructed containing the variables *age*, *household income*, and *socio-economic status*. These auxiliary variables were selected because of their correlation with the target variables and the response probabilities. Table 3.1.2 shows that weighting improved the estimates somewhat. They were closer to the true OViN values. Still, substantial biases remained.

Table 3.1.3

Estimates (unweighted and weighted) for main activity (%)

Main activity	Panel	Weighted	OViN
Housewife/man	11.9	12.2	12.5
Pensioner	16.8	17.8	14.7
At school/student	6.1	10.6	9.8
Disabled	2.4	2.8	2.8
Unemployed	1.9	2.1	2.3
Employed	59.2	52.4	56.1

Table 3.1.3 shows the results for the target variable *main activity*. Pensioners and the employed are over-represented, and students are under-represented. The effects of weighting are mixed. The estimates for some categories (housewife/man, school/student, disabled, unemployed) improve. The weighted estimates are closer to the true value. The opposite effect can be observed for the pensioners and the employed: the weighted estimates are further away from the true value.

3.2 Conclusion

The web panel pilot showed that this mode of data collection can only be used for a limited number of surveys of Statistics Netherlands. Not every CAPI survey can be moved to the web. For example, some survey questionnaires are simply too long. Moreover, drawing samples from the respondents of previous CAPI and CATI surveys leads to panels with a serious lack of representativity. Also, there are substantial costs involved. Particularly, recruitment costs are high. There are also other costs, like yearly maintenance costs (for example, for keeping the panel representative), and costs per survey. This led to decision to postpone the introduction of a web panel for the time being.

References

Baker, R., Blumberg, S.J., Brick, J.M., Couper, M.P., Courtright, M., Dennis, J.M., Dillman, D., Frankel,, M.R., Garland, P.,Groves, R.M., Kennedy, C., Krosnick, J., Lavrakas, P.J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R.K. & Zahs, D. (2010), "Research Synthesis: AAPOR Report on online panels", *Public Opinion Quarterly*, 74, pp. 711–781.

Bethlehem, J.G. (2010), "Selection Bias in Web Surveys", *International Statistical Review*, 78, pp. 161-188.

Bethlehem, J.G. (2013), *De kwaliteit van internetpeilingen*, Inaugural Address, Leiden University: Leiden.

Bethlehem, J.G. & Biffignandi, S. (2012), *Handbook of web surveys*, Hoboken: John Wiley & Sons.

Bethlehem, J.G., Cobben, F. & Schouten, B. (2011), *Handbook of nonresponse in household surveys*, Hoboken: John Wiley & Sons.

Bronzwaer, S. (2012), "Infiltranten probeerden de peilingen van Maurice de Hond te manipuleren", *NRC*, 13 September 2012.

Cook, C., Heath, F. & Thompson, R.L. (2000), "A meta-analysis of response rates in web- or internet-based surveys", *Education and Psychological Measurement*, 60, pp. 821-836.

Kaplowitz, M.D., Hadlock, T.D. & Levine, R. (2004), "A comparison of web and mail survey response rates", *Public Opinion Quarterly*, 68, pp. 94-101.

Knowledge Networks (2012), *KnowledgePanel design summary*, www.knowledgenetworks.com.

Leenheer, J. & Scherpenzeel, A.C. (2013), "Does it pay off to include non-internet households in an internet panel?", *International Journal of Internet Science*, 8, pp. 17-29.

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I. & Vehovar, V. (2008), "Web surveys versus other survey modes – A meta-analysis comparing response rates", *International Journal of Market Research*, 50, pp. 79-104.

Lynn, P. (2013), "Issues of coverage and sampling in web surveys for the general population: an overview". Paper presented at the opening conference of the NCRM Network of Methodological Innovation, Web Surveys for the general population: how, why and when?

Särndal, C.E. & Lundström, S. (2005), *Estimation in surveys with nonresponse*, Chichester: John Wiley & Sons.

Scherpenzeel, A. (2008), "An Online Panel as a Platform for Multi-Disciplinary Research", In: I. Stoop & M. Wittenberg(Eds.), *Access panels and online research, panacea or pitfall?*, Aksant: Amsterdam, pp. 101-106.

Schouten, B., Cobben, F. & Bethlehem, J.G. (2009), “Indicators for the representativeness of survey response”, *Survey Methodology*, 35, Vol 1, pp. 101-113.

Scherpenzeel, A. & Schouten, B. (2011), “LISS Panel R-indicator: representativity in different stages of recruitment and participation of an internet panel”, Paper presented at the 22nd International Workshop on Household Survey Nonresponse, Bilbao, Spain.