Web Panels for Official Statistics?

Jelke Bethlehem and Fannie Cobben
Statistics Netherlands, The Hague, The Netherlands
Corresponding author: Jelke Bethlehem, e-mail: jbtm@cbs.nl

Abstract

Can a web panel be used as a reliable and accurate data collection instrument in official statistics? This is the question the paper attempts to answer. A web panel can be used for either longitudinal or cross-sectional studies. Depending on the type of panel, a number of choices for the implementation need to be made. These involve decisions on survey topics and the questionnaire, the recruitment strategy, maintenance of the panel and how to deal with nonresponse. In this paper, these methodological issues are discussed in more detail.

Keywords: recruitment, nonresponse, attrition, maintenance

1. Introduction

National statistical institutes face an ever growing demand for statistical information about society. The way they collect this information changes over time. Changes in survey methodology and survey technology may be caused by new developments in computer technology (like laptops, the internet, tablets and smart phones), but also by new challenges in society (like increasing nonresponse rates, decreasing budgets, or demands for reducing the response burden).

It is the task of National Statistical Institutes (NSI's) to produce reliable and accurate statistics. Traditionally, they conducted face-to-face or telephone surveys to collect the data for these statistics. This is an expensive way of survey 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.

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 inexpensive 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 sample selection procedures are required. It is just a matter of sending an email to the panel members. No interviewers are involved, 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 that were conducted in 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.

Panels can be used in two different ways. The first one is for *longitudinal research*, in which the same set of variables is measured for the same group of individuals at different points in the time. The focus of research is on measuring change. The second way to use a panel is for *cross-sectional research*. The panel is used as a sampling frame for specific surveys that may address different topics, and thus measure

different variables. Also, samples may be selected from different groups (elderly, high-educated, etc).

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).

In this paper we explore the possible use of web panels for official statistics. Can web panels produce accurate and reliable statistics? Can web panels replace CAPI surveys? This paper attempts to answer these questions by addressing a number of issues, such as under-coverage, recruitment, nonresponse, measurement errors and maintenance.

2. Under-coverage

A web survey may suffer from under-coverage because the target population of a survey is usually much wider than just persons with internet access. According to data of Eurostat, the statistical office of the European Union, 73% of the households in the EU had access to internet in 2011. There were large variations between countries. The countries with the highest percentages of internet access were The Netherlands (94%), Iceland (93%) and Norway (92%). Internet access was lowest in Bulgaria (45%), Romania (47%) and Greece (50%).

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) show that the bias of the sample mean (as an estimator of the population mean) due to under-coverage is determined by two factors. The first factor is the relative size of the proportion of people without internet. The bias decreases as internet coverage increases. The second factor is the contrast, i.e. the difference between the means of the target variable for those with and without internet. The larger the contrast, the larger the bias.

Since internet coverage is steadily increasing, the first factor is decreasing. It is not clear, however, whether the contrast also decreases. To the contrary, it is not unlikely that the (small) remaining group of people without internet will be more and more different from the rest of the population. As a result, some bias may still remain.

One way to solve the under-coverage problem is to provide free internet access to sample persons without it. This approach was implemented in the KnowledgePanel and the LISS panel. This also raises new questions. Is this possible if the under-coverage is large? And what about possible measurement errors caused by those not familiar with the internet?

A different solution of the under-coverage problem, is to extend the web panel into a mixed-mode panel, i.e. maintaining a group of panel members without internet and approaching them for specific surveys in a different mode 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.

3. Recruitment

To set up a web panel that allows for reliable and accurate statistical inference about a general population, a probability sample must be selected. This is not straightforward, because there is no proper sampling frame. 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, or not. 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 survey researcher is not in control of the selection process. Each person has an unknown selection probability, which makes it impossible to construct unbiased estimators. Another problem is that also people from outside the target population can become panel members. People can also have multiple memberships. There are even examples of individuals or groups attempting to manipulate the outcomes of specific surveys.

Bethlehem & Biffignandi (2012) show that the bias of the self-selection sample mean increases if (1) the correlation between the values of the survey variable and the participation probabilities is stronger, (2) the standard deviation of the participation probabilities is larger, and (3) the mean of all participation probabilities is smaller.

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 should be a list of email addresses of all members of 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. One way out is to use another mode of recruitment. For example, Statistics Netherlands selects samples for web surveys from the population register. Selected persons receive a letter with an internet address and a unique login code. Since the response rate is often low (not higher than 40%), nonrespondents are re-approached by telephone (CATI) or face-to-face (CAPI).

Another possible recruitment approach is to invite respondents of another (CAPI or CATI) survey to participate in a web panel. It has the advantage that no new sample needs to be selected and not a new fieldwork operation is required. However, one has to keep in mind that only survey respondents are approached. This may introduce an extra source of selectivity.

4. Nonresponse

Nonresponse occurs in two phases of the 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 panel requires substantial commitment and effort of respondents. The specific survey nonresponse is often low as the invitation to participate is a consequence of agreeing to be a panel member. Causes of nonresponse are not at home, not interested in the specific topic, and not able (e.g. due to illness). Nonresponse need not be permanent. After skipping one of the specific surveys, a panel member may decide to participate again in a subsequent survey.

Attrition is a specific type of survey nonresponse. People get tired of completing the specific survey questionnaires and decide to stop their cooperation. Once they stop, they will never start again.

The problem of nonresponse is that it may be selective. This may cause estimates to be biased. The bias is large if the correlation between response probabilities and the survey variable is strong, the average response probability is low, and the variation of the response probabilities is large. To get some idea of the lack of representativity of the survey response, Schouten, Cobben & Bethlehem (2009) propose to compute the R-indicator. 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 in case of complete representativity. The closer the value of R is to 0, the larger the lack of representativity is. See Scherpenzeel

& Schouten for an example of the use of the R-indicator in a web panel.

To avoid drawing wrong conclusions, some kind of correction must be carried out. Usually, an adjustment weighting technique is applied. A vital ingredient of weighting is the availability of a set of proper auxiliary variables. These variables must have been measured in the panel, and moreover their population distribution must be known.

Weighting adjustment is only effective if two conditions are satisfied: (1) the auxiliary variables must be correlated with response behaviour, and (2) the auxiliary variables must be correlated with the target variables. The set of available auxiliary variables is often limited. If it is not possible to link the survey data file to some administrative data source, often only demographic variables like gender, age, marital status and region are available. One may doubt if these variables have sufficient explanatory power to predict response behaviour or survey variables.

It is wise to conduct weighting in two steps, to correct for recruitment nonresponse and to correct for specific survey nonresponse. These may be different mechanisms requiring different weighting variables and different weighting models.

For the recruitment phase, if relevant auxiliary variables are not available, one might consider conducting a *reference survey* to measure them. This survey should be based on a (small) probability sample. Data collection takes place in a mode different from the web, e.g. CAPI or CATI . Under the assumption of no non-response, or ignorable non-response, this reference survey produces unbiased estimates of the population distribution of the auxiliary variables. Bethlehem (2010) shows that a reference survey can indeed help to reduce a bias, but at the cost of an increased variance of the estimates.

The effects of weighting adjustment of a specific survey can be more effective, because there are more auxiliary variables available. Typically, all members of a web panel completed a so-called profile survey when they were recruited. This profile information is recorded for all panel members. Therefore, all profile variables can be used for weighting adjustment of a specific survey.

5. Measurement errors

Traditionally, many surveys in official statistics are CAPI or CATI surveys. They are used because response rates are high and data quality is good. What would change if these surveys were to be replaced by a web panel? This section focuses on measurement errors and the effects they have on quality.

From a cognitive point of view, answering questions is not an easy task. Schwarz et al. (2008) describe the steps the respondents have to go through: (1) understanding the question, (2) retrieving the required information from memory, (3) translating the information in the proper answer format, and (4) deciding whether to give the answer or not. Several things can go wrong. This may particularly be a problem for web surveys, because there are no interviewers. So they cannot motivate respondents, answer questions for clarification, provide additional information and remove causes for misunderstanding. Respondents are on their own.

Researchers should realize that respondents are usually not interested in the topic of the survey. Therefore participation is not important for them. Krug (2006) describes how people read websites. Many of his points also apply to web survey questionnaires:

- Respondents do not read the text on the screen. They just scan it looking for words or phrases that catch the eye.
- Respondents know there is no penalty for giving wrong answers.

- Respondents do not read the introductory text explaining how the questionnaire works. They just muddle through and try to reach the end.

All this leads to a phenomenon called *satisficing*. Respondents do not select the optimal (true) answer to the question, but the first reasonable answer. Satisficing can take many forms:

- Preference for an answer early in the list (*primacy effect*)
- Agreeing with statements, regardless of their content (*acquiescence*)
- Preference for an answer to keep everything as it is.
- Preference for the neutral middle option.
- Select all answers for a grid question in the same column (*straight-lining*).
- Select the option *don't know* if it is offered as a possible answer.
- Just check a few answers for a check-all-that-apply question.

All these phenomena can lead to incorrect or incomplete answers, and this has a negative effect on the quality of the outcomes.

It should be noted that there are some aspects in which web surveys are better than CAPI and CATI surveys. One of these aspects is sensitive questions. Respondent tend to give socially desirable answers in CAPI and CATI surveys, while their answers are more honest in web surveys.

An additional problem of web surveys is also that it is difficult to control how the questionnaire is displayed on the screen. This is caused by using different browsers, different versions of the same browser, and also by different hardware (desktop, laptop, netbook, tablet, smart phone, etc.).

Research in the area of measurement errors in web surveys is fragmented and often based on small experiments with small groups of students. It is unclear to what extent the results of these experiments can be generalized. More research is required on the magnitude and impact of measurement errors in web panels for the general population.

6. Maintenance

It is important to keep the composition of the panel stable over time. Only then can changes over time be attributed to real changes in society and not to changes in the panel. In this section, three important aspects of web panel maintenance are discussed: specific survey frequency, panel refreshment, and the maximum duration in the panel.

There is a relationship between the accuracy of estimates of change and the frequency of the specific surveys. On the one hand, a high survey frequency increases the (perceived) response burden and therefore may lead to nonresponse. This increases the variance and the bias of estimates. On the other hand, a too low (perceived) response burden may have a negative effect on engagement of panel members.

A web panel may become less representative due to attrition. This makes it important to refresh the panel at certain times. The question is how to do this properly? At first sight, one could think of adding a fresh random sample from the population to the web panel. However, this does not improve the representativity. Those with the highest attrition probabilities remain under-represented.

Ideally, the fresh sample should be selected such that the new members resemble the members that have disappeared due to attrition. Schouten, Cobben and Bethlehem, (2009) show that the R-indicator can be used to identify groups that contribute most to the lack of representativity. The refreshment sample should focus on getting people from these groups in the panel. One should also realize that due to refreshment not all members in the panel will have the same selection probabilities. This should be taken into account when computing unbiased estimates.

Being in a panel for a long time may have an effect on the behaviour and attitudes of the panel members, and even be the cause of a bias. For example, persons may learn how to follow the shortest route through a questionnaire. This effect is called *panel conditioning*. Panel conditioning, may be avoided by restricting panel membership to a specific time period. The maximum time period depends on frequency of the specific surveys, the length of the surveys and also on the variation in survey topics.

7. Conclusions

Coming back to the question if a web panel can be a useful tool for data collection in official statistics, the answer is a careful yes, but only if some conditions are satisfied:

- The problem of under-coverage must be taken care of, either by providing internet to those without it, or by implementing a mixed-mode design.
- Panel members must be recruited by means of probability sampling. Self-selection is out of the question.
- Nonresponse (both during recruitment and in specific surveys) must be prevented as much as possible. Ample relevant auxiliary information must be available to correct for nonresponse bias.
- Sufficient action must be taken to avoid measurement errors as much as possible.
- The panel must be refreshed a certain points in time in order to correct it for the lack of representativity.

One other conclusion is that setting up and maintaining a web panel for official statistics is more complex than maybe anticipated at first sight. It will require more time and money than expected. It is also clear that there are several unsolved problems. This calls for more methodological research.

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. & Biffignandi, S. (2012), *Handbook of Web Surveys*. John Wiley & Sons, Hoboken, NJ

Knowledge Networks (2012), *KnowledgePanel Design Summary*, www.knowledgenetworks.com

Krug, S. (2006), *Don't Make Me Think! A Common Sense Approach to Web Usability, Second Edition.* New Riders, Berkeley, California, USA.

Scherpenzeel, A. (2008), An Online Panel as a Platform for Multi-Disciplinary Research. In: Stoop, I. & Wittenberg, M. (Eds.), *Access Panels and Online Research, Panacea or Pitfall?* Aksant, Amsterdam, The Netherlands, pp. 101-106.

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 the 22nd International Workshop on Household Survey Nonresponse, Bilbao, Spain

Schouten, B., Cobben, F. & Bethlehem, J.G. (2009), Indicators for the Representativeness of Survey Response. *Survey Methodology* 35, pp. 101-113.

Schwarz, N., Knäuper, B., Oyserman, D. & Stich, C. (2008), The Psychology of Asking Questions. In: De Leeuw, E.D., Hox, J.J. & Dillman, D.A. (Eds.), *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates, New York / London, pp 18-34.