Web panels for official statistics

13

Fannie Cobben and Jelke Bethlehem

Discussion paper (201307)

Statistics Netherlands

The Hague/Heerlen, 2013

Explanation of symbols

.	data not available
*	provisional figure
**	revised provisional figure (but not definite)
x	publication prohibited (confidential figure)
–	nil
–	(between two figures) inclusive
0 (0.0)	less than half of unit concerned
empty cell	not applicable
2012–2013	2012 to 2013 inclusive
2012/2013	average for 2012 up to and including 2013
2012/'13	crop year, financial year, school year etc. beginning in 2012 and ending in 2013
2010/'11– 2012/'13	crop year, financial year, etc. 2010/'11 to 2012/'13 inclusive

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

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands
Grafimedia

Cover
Teldesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form:
www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

60083201307 X-10

Web panels for official statistics

Fannie Cobben and Jelke Bethlehem

***Abstract**: New developments in computer technology, but also new challenges in society like increasing nonresponse rates, decreasing budgets, or demands for reducing the response burden, may lead to changes in survey methodology for official statistics. In this paper, the use of a Web panel for high quality statistics about the general population is explored. Such a panel needs to be probability based. 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 questionnaires, the recruitment strategy, maintenance of the panel and how to deal with the various types of nonresponse. In this paper, these methodological issues are discussed in more detail.*

Even if these issues are solved in an acceptable way, then a Web panel will still not be able to replace the traditional survey modes CAPI and CATI. It is, however, possible to identify topics / questions that can be asked in a Web panel without a loss of quality. Maybe even of better quality. In addition, the speed and longitudinal aspect of a Web panel enable the collection of more timely and hence topical information and monitor (changes in) the behaviour of respondents.

The paper ends with an overview of the issues and corresponding suggestions that need further examination to identify the most desirable Web panel practices for official statistics.

Keywords: Longitudinal panel, cross-sectional panel, recruitment nonresponse, attrition, weighting adjustment, panel maintenance, incentives, panel refreshment.

1. Introduction

National statistical institutes have to satisfy an ever growing demand for statistical information in society. The way they attempt to accomplish this changes over time. New developments in computer technology may lead to changes in survey methodology for official statistics, as do new challenges in society like increasing nonresponse rates, decreasing budgets, or demands for reducing the response burden.

National statistical institutes have to produce reliable and accurate statistics. Traditionally, they conduct 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 necessary in order to obtain high quality data. Nowadays, budget constraints cause statistical institutes in many countries to look for less expensive ways of data collection while maintaining data quality.

A Web panel seems a promising alternative. Online data collection has become increasingly popular, particularly in the world of market research. This is not surprising as this is a simple, fast and inexpensive way to collect a lot of data. The largest advantage of a Web panel is the speed with which data can be collected over the Internet, combined with the longitudinal aspect of a panel design. This allows for topical surveys that monitor the behaviour of respondents over time. For this reason, opinions on societal issues and politics are popular topics for panel surveys. Obtaining insight in these actualities is much more complicated in cross-sectional survey research. Also measurements of change can be estimated more accurately than in cross-sectional research. This is interesting for following trends in for example unemployment, health situation or financial behaviour.

At first sight, a Web panel is just another mode of data collection. Questions are not asked face-to-face or by telephone, but over the Internet. However, there are a number of methodological aspects that need to be dealt with in order to use Web panels for official statistics:

- *Recruitment*. Sometimes, researchers rely on self-selection of respondents. Unfortunately, this approach is known to produce biased estimates. Participating in a panel asks more commitment from respondents than participating in a one-time survey, therefore recruitment for a panel is more difficult. In addition, nonresponse in Web surveys is too high to solely use Web for the panel recruitment.

- *Maintenance.* How often should measurements in a panel be made? And how long can members participate in the panel? After taking many surveys, there is a risk of professional respondents. That means that taking many surveys influenced the attitudes and behaviour of respondents. This leads to bias and should be prevented.

- *Nonresponse*. Almost every survey suffers from nonresponse. In a panel study, nonresponse can occur in the recruitment phase. It can also occur in the waves of a panel study. It usually has a monotone pattern: the number of respondents

decreases for each subsequent wave. Once individuals stop responding, they are lost for the panel. This type of nonresponse is usually called attrition. Wave nonresponse may also be caused by the specific topic of the panel survey. Unfortunately, low response rates increase the bias of estimators.

- *Undercoverage*. Since not everyone in the population has access to the Internet, portions of the population are excluded from the panel. This may lead to biased estimates.
- *Measurement errors*. Interviewer-assisted surveys like CAPI and CATI surveys produce high quality data. However, interviewer assistance is missing for web surveys, and this may lead to measurement errors. On the other hand, the risk of social desirable answers is less for web surveys, leading to less measurement errors than in interviewer-assisted surveys.

These methodological issues are discussed in more detail in this paper. It will be shown that undercoverage, self-selection and nonresponse may cause estimators to be biased. One may wonder whether it is possible to correct for such a bias. An overview of some correction techniques is given. It is made clear that applying these techniques is no guarantee for obtaining the same, high quality statistics with a Web panel as can be obtained with traditional survey modes. It is therefore also discussed what other possibilities a Web panel could create for official statistics.

Like in cross-sectional surveys, defining the target population is one of the first actions in designing a Web panel. For official statistics, this is usually the general population. The target population sets restrictions for the method of recruitment of a panel. A panel that is used to produce statistics for the general population needs to be based on a sample that is selected from this population. This is necessary to be able to generalise the findings from the panel to the general population. It is also possible to recruit a Web panel with nonprobability-based methods, like through a banner on a Website. However, with such a recruitment procedure it will be very difficult, if not impossible, to produce statistics for a pre-defined target population.

By applying the principles of probability sampling during recruitment and selection for unique surveys, selection probabilities are known. This enables the computation of unbiased estimates; see the seminal paper of Horvitz & Thompson (1952). Hence, probability sampling is required to recruit persons for a Web panel that needs to be representative for the general population. Two types of panel can be distinguished:

- A longitudinal panel measures the same set of variables for the same individuals at subsequent points in time. In principle, all the members are re-approached each time. Focus of these studies is at measuring change. It is, for example, interesting to follow specific concepts over time, such as unemployment, health status and financial behaviour. Measurements at different points in time on the same set of persons are correlated, thereby reducing the variance of estimates of change.

- A cross-sectional panel is used as a sampling frame for specific surveys. These surveys may address different topics, and therefore may measure different variables. Each of these surveys is in fact a cross-sectional survey.

For longitudinal panels there is only one phase of sampling that takes place during recruitment. Individuals are sampled from the target population, and these individuals are invited to become a member of the panel. For cross-sectional panels there can be two phases of sampling. The first is the same as for longitudinal panels: sampling of panel members in the recruitment phase. The researcher may decide to introduce a second phase of sampling by selecting a sample from the panel for a specific cross-sectional survey. One reason may be that a small sample is sufficient for the purpose of the researcher. Another reason could be to spread the response burden by using different samples for different surveys. And a third reason could be to select the sample from a special group in the population (for example voters for a specific party, smokers, freelancers, school teachers, etc). Characterics that are known from previous waves can be used for this purpose, thus allowing for samples of persons with very specific characteristics that otherwise would not have been available for sampling.

Of course, the two different types of panel can be combined into one panel. A set of questions to measure the change in behaviour is asked in every unique survey, and in addition specific questions are added. With this design there is a risk of order-effects, to avoid these the longitudinal questions should be placed at the beginning of the questionnaire. The type of panel sets restrictions for the design. For example, when the objective is to monitor changes in behaviour it is important that panelists participate every wave. This becomes less important when the Web panel only serves as a sampling frame.

The LISS Panel (Longitudinal Internet Studies for the Social Sciences) is an example of a cross-sectional panel. It consists of approximately 5,000 households. This panel was set up in 2006 by CentERdata, a research institute in The Netherlands. Objective of the panel is to provide a laboratory for the development and testing of new, innovative research techniques, while collecting data for the scientific community.

The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. The initial sample consisted of approximately 10,000 households. Telephone numbers were added to the selected names and addresses. This was only possible for listed numbers. Households with a listed telephone number were contacted by means of CATI. Addresses without a listed number and those who could not be contacted by telephone were visited by the interviewers (CAPI).

During the recruitment phase, individuals in sampled households were first asked to participate in a short interview. They were given an unconditional 10 euro incentive. General background questions about the respondent were asked. At the end of the interview respondents were asked to become a member of the panel. Households without access to Internet, or who were worried that an Internet survey would be too

complicated for them, were told about a simple computer with Internet access that could be installed in their homes for free for the duration of the panel. To demonstrate the use of this computer they were shown a demonstration video.

The Knowledgepanel® is another example of a cross-sectional study panel. It is a probability-based panel in the United States (Knowledgepanel® Design Summary: www.knowledgenetworks.com). The panel consists of approximately 50,000 adult members and 3,000 teenagers with consent of their parents. This panel exists since 1999. Recruitment used to be based on Random Digit Dialing, nowadays an address based sampling frame is used. Persons in selected households are invited to participate. When they do not have access to the Internet, they are provided with a laptop with a paid Internet connection. First, the members of the Knowledgepanel® are profiled (they have to answer questions about demographics as well as attitudinal questions for specific groups). After profiling, they are eligible for selection for specific surveys. Hence, the Knowledgepanel serves as a sampling frame for specific research.

In this paper, a number of methodological issues are described that need to be resolved for the successful implementation of a Web panel for official statistics. In section 2, the recruitment for a Web panel is discussed. In section 3, aspects that are related to the maintenance of a Web panel are discussed. Then, types of nonresponse (section 4) and different adjustment methods for nonresponse (section 5) are presented. The topic of section 6 is measurement errors. This paper ends with a summary and an overview of the methodological issues that need to be resolved in section 7.

2. Recruitment

Recruitment for panel studies is more difficult than for regular, cross-sectional surveys because participating in a panel requires more commitment from respondents. In this section, a number of considerations regarding recruitment for a Web panel are discussed: sampling frame, recruitment method, use of incentives, and undercoverage of persons with no Internet access,

Sampling frame

In the introduction, the distinction between probability-based panels and nonprobability-based panels was made. For nonprobability-based Web panels, the recruitment methods are numerous (e.g. banner recruitment, online advertisement, snowball recruitment, member-get-a-member campaign, recruitment website consolidators, invitations to members of affinity organizations, etc.) but they all have in common the fact that (with very few exceptions) anyone can volunteer to become a panel member. It is not possible for recruiters to know in advance who will see the invitation or to know how many times a given individual might encounter an invitation from a specific panel. As a result, it is impossible for the recruiter to know the probability of selection of any given individual in the entire

population of interest in a survey. Although probability sampling methods may be used to select panel members to be invited to complete a specific survey, this use of probability sampling does not make such surveys probability sample surveys because of the unknown probability of selection inherent in the panel recruitment method.

This type of recruitment is out of the question for official statistics. The main reason is that due to the unknown selection probabilities, it is not possible to rely on sampling theory for statistical inference. But that is not the only problem. By relying on respondents to self-select into the panel, the recruitment procedure is out of control of the researcher. A lot of things can go wrong that are difficult to track and that heavily affect the quality of panel based statistics. Persons from outside the target population can participate. It is even possible to manipulate the outcomes. One person can register for the panel multiple times with different identities[1], or if an incentive is given this can attract persons that are 'in it for the money'. These persons are probably also in other panels in which they can earn money. Answering many surveys may have influenced their behaviour and/or attitudes. It may also lead to a lower quality of their answers because they have become professional in answering surveys. For example, they may be able to identify screening questions and answer them in such a way that minimizes the time needed to complete the survey. These effects are known as *time-in-panel bias* or *panel conditioning*.

Hence, a Web panel for official statistics needs to be probability-based. This is also the first recommendation from the American Association for Public Opionion Research Taskforce on Online Panels (Baker et al. 2010): *Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values.*

For probability-based Web panels, a sampling frame is needed. From this frame, either persons or households can be selected for participation in the Web panel. Surveying households is problematic when it comes to household dynamics (what to do with children that move out, or new partners that move in), partial nonresponse and response burden. It is also more complicated to link records from other data sources, for instance to adjust for nonresponse bias. Therefore, selecting persons is to be preferred. It can be interesting though to collect information on households in stead of just persons. When selecting persons, this could be achieved by interviewing the household to which that person belongs.

Because the mode of datacollection is the Web, it may seem logical to use this mode for recruitment too and therefore to use a sampling frame of e-mail addresses. However, such a sampling frame is generally not available and if it were, it would not cover the general population due to undercoverage of persons without e-mail. A better sampling frame is the population register. In this register, there is no undercoverage of persons without access to the Internet. A sample from the

[1] This can be circumvented by checking on which computer the questionnaire was completed.

population register can be selected with known, non-zero selection probabilities to allow for sound statististical inference. In addition, when such a general population register can be used, some information about the prospective panelists is known and if register information is linked to the selected sample then even more information about prospective panelists is known beforehand. This information can be used to tailor the recruitment approach, for example with the use of adaptive survey design that aims at recruiting a representative subset of panelists. But it can also be used to assess whether the recruitment resulted in a selective pool of panelists, hence providing some insight on the quality of the panel recruitment procedure.

Method of recruitment

Once persons are selected for participation in the web panel, they can be approached in different ways: by web, mail, telephone, face-to-face or a mixture of these modes. Which mode to chose, is often motivated by response rates and costs. It is common to use a different (interviewer-assisted) mode for the panel recruitment. This leads to a higher recruitment response due to the assistance of an interviewer, but is also more costly than recruitment by mail or Web. Of the interviewer-assisted modes, CAPI is known to have the highest response rate and no undercoverage. CATI cannot be used to recruit persons without a listed number, and response rates are lower than in CAPI, but CATI costs less. It is possible to use a combination of CATI and CAPI to balance costs and recruitment response rate.

The current strategy of Statistics Netherlands for Web survey recruitment is to send a letter with a login to a secured website. The sample for the Web survey is selected from the population register from which either persons or addresses (households) are selected. A couple of reminders are send to nonrespondents to remind them about the request for participation. The response that is obtained with this approach is at most 35%, in addition it is selective with respect to age (elderly persons are underrepresentated), ethnic group (non-natives are underrepresented, especially older immigrants from non-Western countries). Because of the selectivity of the response to Web surveys, after the Web approach Statistics Netherlands employs CATI (for sample elements with a listed, fixed-line telephone number) and CAPI (for sample elements without a listed number). This strategy is mostly motivated by costs, using the cheapest mode first (web) and the most expensive mode last (CAPI).

High recruitment nonresponse causes a risk of nonresponse bias. This bias is persistent in all the waves that follow the recruitment. Wave nonresponse and attrition during subsequent waves will only worsen this bias. These two different types of nonresponse need to be dealt with separately. In section 4 and 5 this is discussed in more detail. The recruitment of a representative pool of respondents is critical for the success of a Web panel. Once a person becomes a panel member and participates in the first number of panel surveys, the probability of drop-out (or: attrition) becomes smaller (Baker et al., 2010). Therefore, it makes sense to put a lot of effort in recruiting new panelists and paying more attention to them for the first couple of surveys.

In the LISS-panel (Scherpenzeel, 2009) households with a listed telephone number were contacted by means of CATI. Addresses without a listed number and those who could not be contacted by telephone were visited by the interviewers (CAPI). During the recruitment phase, individuals in sampled households were first asked to participate in a short interview. General background questions about the respondent were asked. At the end of the interview respondents were asked to become a member of the panel. Households without access to Internet, or who were worried an Internet survey would be too complicated for them, were told about the simple-to-operate computer with Internet access that could be installed in their homes for free for the duration of the panel. To demonstrate the use of this computer they were shown a demonstration video.

Table 2.1 shows the response rates in the subsequent phases of the process. The data are taken from Scherpenzeel and Schouten (2011). 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 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 48%.

Table 2.1 also shows the effects 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 2.1. Response rates in the LISS Panel

Phase	Response
Recruitment contact	91 %
Recruitment interview	75 %
Agree to participate in panel	54 %
Active in panel in 2007	48 %
Active in panel in 2008	41 %
Active in panel in 2009	36 %
Active in panel in 2010	33 %

Another possibility to recruit panelists may be by use of existing surveys, where the recruitment question follows the regular interview. However, this means that the recruitment (non)response is preceeded by the response process of the existing survey, leading to an even lower recruitment response.

The Labour Force Survey of Statistics Netherlands is designed as a rotating panel. Households are selected for the first wave, and after the first wave they are asked whether they want to participate in a follow-up survey. For the first wave, data is collected by a mixed-mode data collection scheme. First, all selected adresses are approached to participate in a Web survey. After a number of reminders, nonrespondents to the Web survey are followed-up by CATI when they can be reached by telephone and by CAPI otherwise. In Table 2.2 the response rates and the

recruitment rates following the different data collection modes in the first wave are displayed.

Table 2.2. Response rates and recruitment rates for the first wave of the Dutch Labour Force Survey in June 2012.

	Response wave 1[2]	Recruitment rate	*Total*
Web	26.2%	60.6%	15.9%
CATI	40.2%	92.3%	37.3%
CAPI	39.3%	85.1%	21.9%
% of sample	55.6%		41.3%

As can be seen in Table 2.2, recruitment following another survey resulted in a recruitment rate of 41.3% of the original sample. Also, due to the lack of an interviewer the recruitment amongst the Web survey respondents is much lower than for CATI and CAPI. The highest recruitment rate is obtained for CAPI. It remains to be seen, however, how many households actually participate in wave 2 of the LFS, which will take place in September 2012. However, a recruitment rate of 41% is not bad. Of course, this situation does not completely translate to the situation of recruitment for a Web panel, since the LFS panel requires only 4 follow-up interviews at a 3-month interval. However it seems worthwile to investigate this option for the Web panel recruitment.

Incentives

Offering a small amount of money or a small gift to selected sample elements is called an *incentive*. This can be done to thank persons for the participation, or to motivate them to participate. Incentives can be either promised: conditional on participation, or pre-paid: given to all sample elements regardless of whether they participate and hence unconditional.

The purpose of an incentive is to increase participation. Also, with an appropriate incentive it is possible to reduce nonresponse bias. When due to the incentive persons participate that would not have participated for other reasons, this results in a more balanced group of respondents. Thus, when used appropriately incentives can increase the representativity of the response.

Scherpenzeel (2009) shows that for the LISS-panel, including a pre-paid incentive with a value of 10 euro with the pre-notification letter increased the recruitment response with approximately 25% compared to no incentive. To prevent respondents from not registering after having agreed to participate in the panel, the time between these two events was shortened and an additional 10 euro incentive was given for the panel registration. This proved to be effective: with these measures 48% of the

[2] The response for CATI and CAPI is determined conditionally on the response to Web in previous fieldwork periods due to the sequential character of the data collection process.

sample registered for the panel compared to 38,5% for the group without an incentive. The composition of the panel was selective with respect to elderly, single households and non-Western immigrants (these are all underrepresented) and voters (overrepresented). However, when comparing the composition of the LISS-panel to a traditional face-to-face survey the differences were small, except for the elderly, non-Internet population which is underrepresented in the panel (Scherpenzeel, 2009).

Laurie and Lynn (2008) provide a review of current practice with respect to incentives in longitudinal surveys. They conclude that unconditional, cash incentives are most effective in increasing the response rate. They report on evidence that suggests that repeatedly giving an incentive has a positive effect on response rates. The effect of incentives on attrition bias is less clear. They have also studied the effect of an increase in the value of the incentive and they find that even a small increase can significantly improve response rates. Usually it is possible to find an optimum for the value of an incentive. There are examples in which a very high incentive was found to be counterproductive (see Groves and Couper, 1998). Little is known, however, on the long term effects of incentives on attrition and attrition bias.

Jäckle and Lynn (2007) look into cumulative effects of incentives on nonresponse and bias. They find that the effects of incentives were larger for a mail survey than for a telephone survey, and significantly reduced attrition but had no effect on attrition bias.

Hansen and Pedersen (2012) describe the efficiency of different recruitment strategies for Web panels. They evaluate the effect of different contact strategy, incentives and the use of *social proof* in reminders ('We have already received a great number of registrations, but we would also like to have your opinion on politics.') on the panel recruitment response rate for a Web panel on political issues and media use. They do not include face-to-face recruitment, nor do they evaluate the response to subsequent panel waves. With respect to the contact strategy, they find that, compared to a text message on a mobile phone and a postal letter, recruitment by telephone yields the highest panel recruitment rate. This is also the most expensive panel recruitment method considered in this research. The incentive (a 'prepaid' chocolate bar) did not result in a higher panel recruitment rate. The social proof in the reminder did increase the panel recruitment rate.

Rao, Kaminska and McCutcheon (2010) describe a mail and telephone mode experiment to analyze the effects of incentives, advance letters and telephone reminders on the panel recruitment response rate. They, surprisingly, find that mail recruitment has a higher response rate than recruitment by telephone and is also more cost effective. They do note, however, that this may be due to the fact that the mail group was approached with a well-designed welcome package and that the telephone group first had to agree to participate in the telephone call before they were send the same welcome packet. This is in line with the general finding in the literature that pre-paid incentives are more effective to increase survey participation

compared to conditional incentives. They find that all the conditions considered have an independent, individual positive effect on the panel recruitment rate. Besides the difference in approximation with respect to the welcome package, like Hansen and Pedersen (2012) they do also not consider the response to subsequent waves of the panel.

Recruitment nonresponse compromises survey quality the most, followed by attrition. The literature cited above implies that it pays off to increase recruitment response by employing pre-paid, cash incentives. However, if the increase in recruitment response is offset by an increase in attrition in later waves, the positive effect of an incentive is more likely to be outweighted by the costs. However, the reported evidence suggests that the effect of an incentive is persistent for subsequent waves. There is, however, not much guidance in the literature on how incentives should be employed to maximise response while minimising bias for the various types of nonresponse in panel studies. Whether an incentive should be given at recruitment only, for every panel survey, to specific groups, depending on the duration of the panel survey, etc. depends on the effect on the recruitment response rate, the attrition rate and the corresponding biases.

Undercoverage

Undercoverage is the phenomenon that not every member of the target population is represented in the sampling frame. This can happen in the recruitment phase for a Web panel for the general population if not each member of this population has access to the Internet.

Undercoverage leads to bias when the characteristics that are related to Internet access are also related to the survey topic. Undercoverage can be prevented by providing persons without Internet with a computer and an Internet connection. In the LISS-panel, households without access to Internet, or who were worried an Internet survey was too complicated for them, were told about a simple computer with Internet access that could be installed in their homes for free for the duration of the panel (SIMPC). To demonstrate the use of this computer they were shown a demonstration video. In the LISS-panel, 93% of the panelists has access to the Internet (hence 7% use the SIMPC), compared to 85% in the Dutch population in 2008; nowadays it is 95% (source: Statistical yearbook 2011, Statistics Netherlands). Hence, the underrepresentation of the population without access to the Internet is not completely solved by providing a computer with Internet access. In addition, there will always be persons that do not want to answer surveys over the web, or for whom that is too difficult. Also the Knowledgepanel® provides persons without Internet access with a laptop that has access to the Internet.

To adjust for the bias due to undercoverage, a weighting procedure can be applied. Different weighting methods are presented in section 5. There is, however, no guarantee that weighting is successful in reducing the bias.

Dealing with recruitment nonresponse

That nonresponse - in whatever form - is a threat to the quality of estimates based on a Web panel will be obvious from the discussion above. Incentives can be used to decrease the nonresponse rate. But the panel design itself also offers opportunities to deal with nonresponse. Especially for wave nonresponse and attrition, a lot is known about respondents from previous waves of the panel or from recruitment. This information can be used to determine the selectivity of the nonresponse or attrition with respect to substantive variables, i.e. the target variables that are typically unknown in cross-sectional survey research. These values are also not known for the recruitment nonrespondents. Therefore, these two types of nonresponse should be handled separately, see also section 5.

Another way to deal with nonresponse is the use of a reference survey, or questions in a reference survey. Recently, Statistics Netherlands re-designed its household statistics. From unimode surveys, all surveys are transformed to a mixed-mode design. Part of the re-design was the introduction of a so-called basic questionnaire. This questionnaire precedes all re-designed household surveys and contains questions about employment situation, education and health. When the same questions are asked in the Web panel, comparing the results to the basic questionnaire from other SN-surveys may provide valuable insights in the representativity of the (recruited) response to the Web panel.

Nonresponse and attrition may be prevented by making the reasons why persons participate more salient to them, following the leverage-saliency theory from Groves et al. (2000). Possible reasons for participating in a panel may be (Baker et al., 2010): a (monetary) incentive, the importance of expressing or registering one's opinion, the entertainment value of taking surveys, the opportunity to find out what other people think (social comparison), the ease of joining and participating (convenience).

It may be possible to identify groups that participate for the same reason, thus allowing for the tailoring of the approximation strategy. This can for instance be done by variation of the content of the pre-notification letter. Identifying these groups can, for example, be done by a follow-up survey in which panelists are asked for which reason they participated.

3. Maintenance

As discussed in the introduction, there are two types of panel studies: a longitudinal panel study that can be used to measure changes and a cross-sectional panel study that is used as a sampling frame, for example for research in specific groups.

For the longitudinal panel, there are a number of questions on specific topics that are asked at regular time intervals. Behaviour of panelists can be monitored over time. For the cross-sectional panel, a sample is selected from the panel members for unique surveys. Characteristics that are known from previous waves can be used for this purpose, allowing for samples of persons with very specific characteristics that

otherwise would not have been available for sampling (an exception may be the use of screener questions in cross-sectional surveys based on which persons are selected for follow-up research).

For longitudinal research purposes, it is important that panelists, once recruited, remain active in the panel and participate in (almost) every wave. It should be prevented that persons drop out of the panel. When the Web panel serves as a sampling frame, the rules regarding panel membership may be less strict. If a Web panel is used for both longitudinal and cross-sectional studies, the maintenance of the panel should be guided by the most restrictive panel type, being the longitudinal panel.

In this section, three important aspects with respect to the maintenance of a Web panel are discussed: frequency of measurements, refreshment of the panel, maximum duration in the panel.

Frequency of measurements

In a panel study, repeated measurements can be made on the same group of persons at different points in time. This allows for longitudinal analyses such as the measurement of change. It is, for example, interesting to follow specific concepts over time, such as unemployment, health status and financial behaviour. Measurements at different points in time on the same set of persons are correlated, thereby reducing the variance of estimates of change.

There is a dependency between the accuracy of estimates of change and the frequency of measurements. This is caused by panel attrition. A high number of measurements increases the response burden and will lead to more attrition. This increases the variance of the estimates, and also the bias. A high correlation decreases the variance. Depending on the correlation, and the attrition rate, the frequency of the measurements can be determined. This also depends on the response burden.

The (perceived) response burden is determined by topic interest, duration of the survey, the frequency of the surveys and the rotation scheme. A response burden that is too high will lead to wave nonresponse and attrition. But a (perceived) response burden that is too low will decrease the engagement of panelists. For the longitudinal research topic it should be determined by subject matter experts how often they want to measure certain behaviour. Behaviour that is more subjective to changes, like political preference, should be monitored more frequently than behaviour that changes less often, like employment situation.

To lower the response burden and hence control errors due to nonresponse and attrition, a rotating design is usually chosen. This implies that the panel consists of a number of groups or cohorts. At regular intervals, for example each month, a new group of panellists is recruited and another group reaches the maximum duration and drops out of the panel. This may affect the accuracy of measurements of change. Rotating groups imply that there is less overlap between the measurements, hence a lower correlation and a lower accuracy. For the rotation scheme, it should be

determined how often new groups are recruited, and how long panelists can stay in the panel. This is different from panel refreshment, which will be discussued later. There are few clues in the literature how the optimal rotation scheme can be determined. This depends also on the frequency of the measurements, and the involvement of panelists (which is determined by the response burden) and needs to be evaluated in practice.

The Knowledgepanel® administers three to four surveys a month to panelists. The surveys are kept short, with durations of 5 to 20 minutes at most. For each survey, panelists earn points with a cash value of one dollar. Surveys that take longer than 16 minutes are rewarded with an additional, small incentive. This is done to minimise panel attrition by actively involving panelists at an acceptable level of response burden. In addition, the Knowledgepanel has a so-called Panel Relations program to keep panelists involved. The LISS-panelists are sent one survey each month, which takes about half an hour to complete. Like in the Knowledgepanel® they receive a small amount (€7.50) of money for each survey that they complete.

Attrition is most likely amongst the newest members (Baker et al., 2010). Once persons agree to participate in the panel and have answered some surveys, they are likely to continue doing so. This is because people like to behave in a way that is consistent with their own previous behaviour (Cialdini, 1988). So most of the effort should be directed at recruiting panelists and stimulating them to participate with the first few surveys.

Panel refreshment

Panel refreshment means that a fresh sample is selected and recruited for the panel to compensate for attrition. If it is assumed that the panel after recruitment is representative for the target population, finding replacement for the attriters could prevent the panel to become less representative due to attrition. A refreshment sample can be a simple random sample. This is, however, not a solution to keep the panel representative. It would be wise to replace attriters selectively, hence to draw a selective sample from the population that resembles the group of attriters so that the composition of the panel remains as representative as it was before the attrition. The question arises how this can be done in such a way that it is still possible to calculate the inclusion probabilities.

There are some drawbacks to this approach. For a selective selection of the sample, only characteristics that are available for the entire target population (or sample) can be used. The Missing-At-Random (MAR) assumption needs to hold for the characteristics that are not available for sampling in order to really compensate for the selective attrition. The refreshment sample has to be recruited, leading to (recruitment) nonresponse that is probably also selective. And to fine-tune the selective sampling to the selective attrition, it is not possible to recruit a refreshment sample before the actual attrition occurred. Therefore, from wave to wave there will always be some selectivity due to attrition. This can be dealt with by weighting adjustment methods, as described in section 5.

Another approach could be to recruit possible future panelists and to use characteristcs from the recruitment interview for selective panel refreshment besides the information that is available for the entire sample from registers. This way, the MAR-assumption is more likely to hold. However, this also introduces an additional source of nonresponse.

Selective panel refreshment increases the quality of the Web panel and should be implemented in practice. The R-indicator can be used to identify groups that contribute most to the selectivity of the panel composition (Schouten, Cobben and Bethlehem, 2009).

Maximum duration

When respondents answer many surveys, this may influence their behaviour and attitudes introducing a bias. For example, persons learn how to follow the shortest route through a questionnaire. This effect is known as *panel conditioning*, and will be discussed in more detail in section 6. To avoid panel conditioning, it is common to set a maximum to the time that respondents can be a member of the panel. What the maximum duration should be depends on frequency of the panel surveys, the duration of the survey and also on the variation in survey topics. Research has shown that in probability-based panels the effect of panel conditioning can be controlled by varying the topic of the panel surveys from wave to wave (Dennis, 2001).

A comparison of 19 Dutch online panels in 2006 (Van Ossenbruggen, Vonk and Willems, 2006) showed that most attrition took place within the first few months of the panel membership. It should be noted that all these panels were non-probability, self-selection panels. After 6 months, the response stabilised. In the first year of the panel membership, the response was highest. They compared online panels that were recruited in different ways. The number of survey requests per month was 1 to 2 for 60% of the panels, for 35% of the panels this was even more. The average time needed to complete a survey was somewhere between 10 and 15 minutes. All panels pay for survey participation, on average 1.20 euro per 10 minutes.

4. Nonresponse in Web panels

Type of nonresponse

In a panel, there are four types of nonresponse: item nonresponse, unit nonresponse during recruitment, unit nonresponse in specific waves and drop-out. In this paper, item nonresponse is not discussed.

Unit nonresponse during recruitment is referred to as *recruitment nonresponse*. In case of a longitudinal panel, people may drop-out of the panel, and never come back again. This type of nonresponse is called *attrition*. In case of a cross-sectional panel, people may not respond in a specific survey. This is called *survey nonresponse*. Survey nonresponse need not be permanent. It can happen that a person decides to

participate in a subsequent survey, for example if he or she is more interested in the topic of the survey.

Basically, nonresponse problems in a panel do not differ much from nonresponse problems in cross-sectional surveys. However, in a panel study much more is known on the nonrespondents from the recruitment for the panel and from previous waves or surveys. This information could be used to adjust for nonresponse bias. There are, however, not many published studies that in fact do so.

Usually, nonresponse and attrition do not follow a random pattern. Detecting the effects of panel attrition can be done by correlating the duration of stay in the panel with relevant variables. If there are large effects, a solution would be to treat panel duration as an explanatory variable in analyses and calculate all estimates conditional on duration = 0. Another solution may be panel refreshment (see section 3). This means that a fresh sample is selected and recruited for the panel to compensate for the selective attrition so that the panel is again representative for the target population.

How to ensure a representative response?

When the response probabilities of specific groups in the population are known beforehand, these can be used to make strata in such a way that it is possible to select a sample that – in expectation – results in a balanced response with respect to the characteristics chosen to form the strata. However, this is no guarantee of a representative response with respect to the topic of the survey. Nonrespondents can be similar to respondents for the characteristics based on which strata are formed, but can still be different when it comes to the survey topic. This measure will result in a higher number of respondents in typically underrepresented groups, but any underlying reason for nonresponse – that may be related to the survey topic and hence introduce nonresponse bias – is left untouched. Therefore this measure will simply result in 'more of the same' respondents.

It would be better to take care of the lack of representativity during data collection, for example by adaptive or responsive survey design for the recruitment of a panel. The aim of such designs is to influence the response probabilities of recruited panelists in such a way that the random selection mechanism is preserved and the assumption that nonrespondents resemble respondents on survey topic related characteristics is more likely to hold.

The response rate is an important indicator for the quality of single, cross-sectional surveys, because low response rates present a serious risk of unreliable estimates. See for example Bethlehem et al. (2011). Use of response rates in the context of panels can be less meaningful. If the response is high in one of the surveys from of a panel, but it was low during the recruitment phase, the high wave response rate does not guarantee the bias of the estimates to be small. A good sample from a bad panel does not produce reliable estimates.

There is a lot of literature focusing on the question of the usefulness and meaning of response rates. See, for example, American Association for Public Opinion Research

(AAPOR, 2006). For panels based on probability sample recruitment, the cumulative or multiplicative response rate can be used. This computation takes into account the response rate of the recruitment phase and the response rates of the subsequent surveys. See Huggins and Eyerman (2001), Schlengen et al. (2002), Tourangeau (2003), Office of Management and Budget (2006), Schonlau et al. (2009) and Couper (2007). Nevertheless, even in the case of probability-based web panels, the computation of response rates can be quite a difficult task if a possibly complex procedure has to be taken into account. No rigorous and well-accepted terminology and definition of response rate in web panels exists yet.

Before introducing and describing a number of indicators, the following points are stressed:

- There are different stages in building a web panel. In each stage focus is on different kinds of response rates and indicators;
- Cumulative response rates over consecutive stages can be used to evaluate the performance and quality of studies using probability-based panels;
- Computation of indicators should be restricted to only the active part of the panel. For the computation of response rates over time, the availability of panel members at a given point in time must be taken into account.

Callegaro and DiSogra (2008) propose a systematic framework of concepts and indicators related to response in panels. The basic idea is that the computation of response rates for a probability-based panel has to take into account all steps in the recruitment and maintenance. At each step, different response rates can be computed and each of them provides insight in different aspects of the quality and success of the survey. An example of the response rates for each of the different stages of recruitment in the Dutch LISS-panel is displayed in Table 2.1. A number of indicators are introduced below.

The *recruitment rate RECR* is the response rate in the recruitment phase. It measures success of recruitment.

Persons agreeing to participate in a web panel usually are asked to do a profile survey in which basic demographic information is collected. This information can be used to select specific groups for a survey, or to correct for nonresponse. The response rate of the profile survey is called the *profile rate*, and it is denoted by *PROR*.

Once a panel is in operation, members of the panel are contacted for participation in specific surveys. This is usually done by means of e-mail. Not all e-mails will reach the panel members. For example, some e-mails may end up in spam filters. The *absorption rate* is defined as the fraction of e-mails that is really read. It is an indicator of the success of the e-mail operation and denoted by *ABSR*.

The *completion rate* is the proportion of selected and invited members of the panel who completed a specific web survey. Thus, this rate reflects the success of a specific study. The *completion rate* is denoted by *COMR*.

Break-off occurs when a person starts answering the questions in a survey questionnaire, but does not complete the questionnaire. The questionnaire was started but not finished. The *break-off rate* measures the magnitude of this phenomenon. It is defined as the fraction of break-off cases. It is denoted by *BOR*.

In case of probability-based panels, cumulative response rates can be computed. These indicators measure what happens in the subsequent steps of a survey, from panel recruitment to response in a specific study. These cumulative response rates are obtained by multiplying rates that have been obtained for each step in the process.

Two types of cumulative response rates can be calculated. They are called the cumulative response rate 1 and the cumulative response rate 2.

The cumulative Response Rate 1 (*CUMRR1*) is defined by:

$$CUMRR1 = RECR \times PROR \times COMR$$

This response rate reflects the fraction of cases left over after nonresponse in the recruitment phase, nonresponse in the profile survey, and nonresponse in the specific web survey.

The cumulative response rate 2 introduces a fourth component: the retention rate. The *retention rate* (*RETR*) is the proportion of the original sample that remains in the active panel at the time the sample for the specific survey is drawn. Therefore, cumulative response rate 2 is defined with reference to a specific cohort. For a given cohort, this indicator is obtained by multiplying the cumulative response rate 1 by the retention rate, *cumulative response rate 2* is defined by

$$CUMRR2 = RECR \times PROR \times RETR \times COMR$$

Attrition is defined as the proportion of active panel members that drop out of the panel in a specific time period. The attrition rate is denoted by *ATTR*.

A more detailed description of the various response indicators for web panels is given by Bethlehem and Biffignandi (2012).

5. Weighting adjustment

Weighting adjustment is a family of techniques that attempt to improve the accuracy of survey estimates by using auxiliary information. *Auxiliary information* is defined as a set of variables that have been measured in the survey, and for which information on their population distribution (or complete sample distribution) is available.

By comparing the response distribution of an auxiliary variable with its population (or complete sample) distribution, it can be assessed whether or not the sample is representative for the population (with respect to this variable). If this distribution differs considerably, one must conclude that the sample is selective. To correct this,

adjustment weights are computed. Weights are assigned to all records of observed individuals. Estimates of population characteristics are then computed by using the weighted values instead of the unweighted values.

Weighting adjustment is often used to correct surveys that are affected by nonresponse. An overview of weighting adjustment can be found in Bethlehem and Biffignandi (2012) and Särndal and Lundström (2005).

In this section, various weighting techniques are explored. The focus is on weighting adjustment for cross-sectional study panels, but note that these techniques are equally well applicable for longitudinal study panels.

As presented in the previous section, nonresponse can occur in the different phases of the selection process: during recruitment and during the subsequent surveys of the cross-sectional study panel. This implies that also two corrections are required. A first approach could be to ignore the two phases of nonresponse. This implies that weights are obtained by directly confronting response distributions for auxiliary variables in the survey with its population distribution. This is not the most effective way to conduct adjustment weighting. Weighting in two steps should be preferred. In the first place, recruitment nonresponse may be a different phenomenon than survey nonresponse. Therefore it may require a different model containing different variables. In the second place, there are a lot more auxiliary variables available to correct for the survey nonresponse. For many Web panels, new members conduct a profile survey. They have to complete a questionnaire asking basic demographic questions. All these variables can be used to weight the survey 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.

This section explores the possibility to reduce the bias of web survey estimates. For sake of convenience only weighting adjustment for the recruitment phase are described, but note that weighting adjustment for each separate survey is completely similar.

To keep things simple it is assumed the all individuals in the target population have access to the Internet. So, there are no undercoverage effects. Note that the techniques described here can also be used to adjust for bias due to undercoverage. It is also assumed that a simple random sample has been selected from the population.

In section 5.1, some general concepts in survey sampling and the corresponding notation is introduced. Section 5.2 explores the use of post-stratification, where weights are computed using the distribution of auxiliary variables in the complete

population. Section 5.3 investigates the situation where the population distribution of auxiliary variables is estimated using data from a small, true probability sample (a so-called *reference survey*). Finally, section 5.4 discusses propensity weighting, a weighting technique often applied by commercial market research agencies.

5.1 The nonresponse bias

Let the *target population U* of the panel consist of a set of N individuals, which are labelled 1, 2, ..., N. Associated with each individuals k is an unknown value Y_k of the *target variable*. The set of all values of the target variable is denoted by Y_1, Y_2, …, Y_N. Objective of a specific survey is assumed to be estimation of the population mean

$$\bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k \quad . \tag{1}$$

To construct a panel, a simple random sample of size n is selected without replacement. The sample is represented by the series of indicators a_1, a_2, …, a_N, where the k-th indicator a_k assumes the value 1 if individual k is selected in the sample, and otherwise it assumes the value 0.

In case of simple random sampling without replacement, the sample mean

$$\bar{y} = \frac{1}{n}\sum_{k=1}^{N} a_k Y_k \tag{2}$$

is an unbiased estimator of the population mean. Now suppose there is nonresponse in the recruitment phase. It is assumed that each individual k in the population has a certain, unknown probability ρ_k of response. If individual k is selected in the sample, a random mechanism is activated that results with probability ρ_k in response and with probability 1 - ρ_k in nonresponse. Under this model, a set of response indicators R_1, R_2, …, R_N can be introduced, where R_k = 1 if the corresponding element k responds, and where R_k = 0 otherwise. So, $P(R_k = 1) = \rho_k$, and $P(R_k = 0) = 1 - \rho_k$.

The recruitment phase response only consists of those elements k for which a_k = 1 and R_k = 1. Hence, the number of available individuals is equal to

$$n_R = \sum_{k=1}^{N} a_k R_k \ , \tag{3}$$

The number of nonrespondents is denoted by n_{NR}, where $n = n_R + n_{NR}$. The values of the target variable only become available for the n_R responding individuals. The mean of these values is denoted by

$$\bar{y}_R = \frac{1}{n_R}\sum_{k=1}^{N} a_k R_k Y_k \ . \tag{4}$$

Bethlehem (2009) shows that the expected value of the response mean is approximately equal to

$$E(\bar{y}_R) \approx \frac{1}{N}\sum_{k=1}^{N}\frac{\rho_k}{\bar{\rho}}Y_k \tag{5}$$

where

$$\bar{\rho} = \frac{1}{N}\sum_{k=1}^{N}\rho_k \tag{6}$$

is the mean of all response probabilities in the population. Expression (5) shows that, generally, the expected value of the response mean is unequal to the population mean to be estimated. Therefore, this estimator is biased. This bias is approximately equal to

$$B(\bar{y}_R) = \frac{R_{\rho Y} S_\rho S_Y}{\bar{\rho}}, \tag{7}$$

where $R_{\rho Y}$ is the correlation coefficient between the values of the target variable and the response probabilities, S_ρ is the standard deviation of the response probabilities, and S_Y is the standard deviation of the variable Y. From this expression of the bias a the following conclusions can be drawn:

- The bias vanishes if there is no relationship between the target variable of the survey and the response behavior ($R_{\rho Y} = 0$). The stronger the correlation between target variable and response behavior, the larger the bias will be.
- The bias vanishes if all response probabilities are equal ($S_\rho = 0$). Indeed, in this situation the nonresponse is not selective. It just reduces the sample size. The more the values of the response probabilities vary, the larger the bias will be.
- The magnitude of the bias increases as the mean of the response probabilities decreases. Translated in practical terms, this means that lower response rates will lead to larger biases.

5.2 Post-stratification

Post-stratification is a well-known and often used weighting technique; see e.g. Cochran (1977) or Bethlehem (2002). To carry out post-stratification, qualitative auxiliary variables are needed. Here, only one such variable is considered. The extension to more variables is not essentially different. Suppose there is an auxiliary variable X having L categories. So it divides the target population into L strata. The strata are denoted by the subsets U_1, U_2, ..., U_L of the population U. The number of target population elements in stratum U_h is denoted by N_h, for h = 1, 2, ..., L. The population size N is equal to $N = N_1 + N_2 + ... + N_L$. This is the population information assumed to be available.

Suppose a simple random sample is selected from the population, and n_R persons respond. If $n_{R,h}$ denotes the number of respondents in stratum h, then $n_R = n_{R,1} + n_{R,2} + ... + n_{R,L}$.

Post-stratification assigns identical adjustment weights to all elements in the same stratum. The weight w_k for an element k in stratum h is equal to

$$w_k = \frac{N_h / N}{n_{R,h} / n_R} \tag{8}$$

The response mean

$$\bar{y}_R = \frac{1}{n_R} \sum_{k=1}^{N} a_k R_k Y_k \tag{9}$$

is now replaced by the weighted sample mean

$$\bar{y}_{R,PS} = \frac{1}{n_R} \sum_{k=1}^{N} w_k a_k R_k Y_k \tag{10}$$

Substituting the weights and working out this expression leads to the post-stratification estimator

$$\bar{y}_{R,PS} = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{y}_R^{(h)} = \sum_{h=1}^{L} W_h \bar{y}_R^{(h)}, \tag{11}$$

where $\bar{y}_R^{(h)}$ is the response mean in stratum h and $W_h = N_h / N$ is the relative size of stratum h. The bias of this estimator is equal to

$$B(\bar{y}_{R,PS}) = \sum_{h=1}^{L} W_h B(\bar{y}_R^{(h)}) = \sum_{h=1}^{L} W_h \frac{R_{\rho Y}^{(h)} S_\rho^{(h)} S_Y^{(h)}}{\bar{\rho}^{(h)}}, \tag{12}$$

where $R_{\rho Y}^{(h)}$ is the correlation between Y and ρ in stratum h, $S_\rho^{(h)}$ the standard deviation of the response probabilities in stratum h, and $S_Y^{(h)}$ is the standard deviation of the target variable in stratum h.

Expression (4.5) shows that the bias of the post-stratification estimator is a weighted sum of the biases of the stratum estimators. The bias will be small if there is a strong relationship between the target variable Y and the stratification variable X. The variation in the values of Y manifests itself between strata but not within strata. In other words, the strata are homogeneous with respect to the target variable. In nonresponse correction terminology, this situation comes down to Missing-At-Random (MAR).

The bias of the estimator will also be small if the variation of the response probabilities is small within strata. This implies that X must be a strong relationship between the auxiliary variable and the response probability.

The conclusion is that application of post-stratification will successfully reduce the bias of the estimator if proper auxiliary variables can be found. Such variables should satisfy three conditions:

- They have to be known for at least the survey respondents (this can be achieved by measurement in the survey or by linkage to a register);
- Their population distribution (N_1 ,N_2, ..., N_L) must be known;

- They must be strongly correlated with all target variables;
- They must be strongly correlated with the response behaviour.

Unfortunately, such variables are not very often available, or there is only a weak correlation.

Post-stratification is only a simple and straightforward weighting technique. More advances weighting adjustment techniques are described in Bethlehem (2002) and Särndal & Lundström (2005). They are denoted by *generalized regression estimation* (*linear weighting*) and *raking ratio estimation* (*multiplicative weighting*).

It should be noted that all these weighting techniques will only be successful if the Missing-At-Random (MAR) assumption applies.

5.3 Weighting adjustment with a reference survey

The previous section showed that post-stratification can be an effective correction technique provided auxiliary variables are available that have a strong correlation with the target variables of the survey and with the response behaviour. If such variables are not available, one might consider conducting a *reference survey*. This reference survey is based on a probability sample, where data collection takes place with a mode leading to high response rates and little bias, e.g. CAPI (Computer Assisted Personal Interviewing, with laptops) or CATI (Computer Assisted Telephone Interviewing). CAPI and CATI survey tend to have high response rates. They can be used to produce accurate estimates of population distributions of auxiliary variables. These estimated distributions can be used as benchmarks in weighting adjustment techniques.

The reference survey approach has been applied by several market research organisations; see e.g. Börsch-Supan et al. (2004) and Duffy et al. (2005). They used the reference survey approach to reduce the bias caused by self-selection of respondents (instead of a probability sample).

To explore this, it is assumed that one qualitative auxiliary variable X is observed both in the recruitment of the Web panel and in the reference survey, and that this variable has a strong correlation with the target variable Y of the survey. Then a form of post-stratification can be applied where the stratum means are estimated using Web panel data and the stratum weights are estimated using the reference survey data. This leads to the post-stratification estimator

$$\bar{y}_{R,RS} = \sum_{h=1}^{L} \frac{m_h}{m} \bar{y}_R^{(h)} \tag{13}$$

where $\bar{y}_R^{(h)}$ is the Web panel based estimate for the mean of stratum h of the population (for $h = 1, 2, ..., L$) and m_h / m is the relative sample size in stratum h for the reference sample (for $h = 1, 2, ..., L$). Under the conditions described above the quantity m_h / m is an unbiased estimate of $W_h = N_h / N$.

It can be shown, see e.g. Bethlehem & Biffignandi (2012), that the expected value of estimator (13) is identical to that of the post-stratification estimator in section 5.4.

So, use of a reference survey with the proper auxiliary variables can substantially reduce the bias of Web panel estimates.

An interesting aspect of the reference survey approach is that any variable can be used for adjustment weighting as long as it is measured both in the reference surveys and in the Web panel. For example, some market research organisations use ‘webographics’ or ‘psychographic’ variables that divide the population in 'mentality groups'. See Schonlau et al. (2004) for more details about the use of such variables.

It should be noted that use of the estimated population distribution will increase the variance of the estimator. The increase in variance depends on the sample size of the reference survey: the smaller the sample size, the larger the variance. So, using a reference survey comes down to reducing the bias at the cost of increasing the variance.

5.4 Propensity weighting

Propensity weighting is used by several market research organisations to correct for a possible bias in their web surveys. Examples can be found in Börsch-Supan et al. (2004) and Duffy et al. (2005). The original idea behind propensity weighting goes back to Rosenbaum & Rubin (1983, 1984). They developed a technique for comparing two populations. They attempt to make the two populations comparable by simultaneously controlling for all variables that were thought to explain the differences. In the case of a Web panel, there are also two populations: those who participate in the Web panel (if asked), and those who will not participate. Under specific assumptions, propensity scores can also be used to control for measurement effects; in that case the two populations are respondents to the Web panel, and respondents to a reference survey. It is assumed that the information used to calculate propensity scores is able to explain the difference in selection between the two modes. Then, by applying the propensity score method, the difference in composition of the pool of respondents to both surveys is controlled for and mode effects due to measurement errors remain and can be adjusted for. The focus in the remainder of this section is on the use of propensity weighting to adjust for nonresponse bias.

Propensity scores are obtained by modelling the variable that indicates whether or not someone participates in the survey. Usually a logistic regression model is used where the indicator variable is the dependent variable and attitudinal variables are the explanatory variables. These attitudinal variables are assumed to explain why someone participates or not. Fitting the logistic regression model comes down to estimating the probability (propensity score) of participating, given the values of the explanatory variables.

Application of propensity weighting assumes some kind of random process determining whether or not someone participates in the web survey. Each element k in the population has a certain, unknown probability ρ_k of participating, for $k = 1, 2, .., N$.

Suppose a simple random sample of size n is selected. The sample can be represented by the series of indicators $a_1, a_2, \ldots, a_N$, where $a_k = 1$ if individual k is in the sample, and $a_k = 0$ otherwise. Nonresponse is denoted by the series of indicators $R_1, R_2, \ldots, R_N$ denote indicator variables, where $R_k = 1$ if person k responds, and $R_k = 0$ otherwise. Consequently, $P(R_k = 1) = \rho_k$, where ρ_k is the probabilities of response.

The propensity score $\rho(X)$ is the conditional probability that a person with observed characteristics X responds, i.e.

$$\rho(X) = P(R = 1 \mid X)$$

It is assumed that within the strata defined by the values of the observed characteristics X, all persons have the same response probability. This is the Missing At Random (MAR) assumption. The propensity score is often modelled using a logit model:

$$log\left(\frac{\rho(X_k)}{1-\rho(X_k)}\right) = \alpha + \beta' X_k$$

The model is fitted using Maximum Likelihood estimation. The estimate of the propensity score $\rho(X_k)$ is denoted by $\hat{\rho}(X_k)$.

To be able to properly estimate the propensity scores, two conditions have to be fulfilled. The first condition is that proper auxiliary variables are available. These are variables that are capable of explaining whether or not someone is willing to participate in the web survey. Variables often used measure general attitudes and behaviour. They are sometimes referred to as 'webographic' or 'psychographic' variables. Schonlau et al. (2004) mention as examples "Do you often feel alone?" and "On how many separate occasions did you watch news programs on TV during the past 30 days?".

The second condition is that the values of the auxiliary variables that are used in the computation of the propenisity scores must be known for the complete sample. This is generally not the case. A possible solution to this problem is to carry out a reference survey. To allow for accurate estimation of the population distribution, the reference survey must be based on a true probability sample from the target population, and the nonresponse mechanism must be MCAR or MAR. Calibration methods can then be used to calculate propensity scores.

Such a reference survey can be small in terms of the number of questions asked. It can be limited to the webographic questions. Preferably, the sample size of the reference survey should be large to allow for accurate estimates. A small sample size results in large standard errors of estimates.

Once response propensities have been estimated, they can be used to reduce a possible response bias. There are two general approaches: response propensity weighting and response propensity stratification.

Response propensity weighting is based on the principle of Horvitz & Thompson (1952) that always an unbiased estimator can be constructed if the selection

probabilities are known. In case of nonresponse, selection depends on both the sample selection mechanism and the response mechanism. The idea is now to adapt the Horvitz-Thompson estimator by including the response probabilities. Suppose the sample is selected by means of a simple random sample. Then the response mean is a biased estimator, but an unbiased estimator would be

$$\bar{y}_{R,HT} = \frac{1}{n}\sum_{k=1}^{N}\frac{a_k R_k Y_k}{\rho_k}. \tag{14}$$

Unfortunately, the response probabilities ρ_k are not known. Now, they are replaced by the estimated response propensities $\hat{\rho}(X_k)$, resulting in the estimator

$$\hat{\bar{y}}_{R,HT} = \frac{1}{n}\sum_{k=1}^{N}\frac{a_k R_k Y_k}{\hat{\rho}_k}. \tag{15}$$

The closer the estimated response propensities are to the original response probabilities, the smaller the bias of the estimator will be.

There are more advanced estimators than the Horvitz-Thompson estimator. One example is the generalized regression estimator. Also this estimator can be improved by including response propensities. For more details, see Bethlehem, Cobben and Schouten (2011).

Response propensity stratification takes advantage of the fact that estimates will not be biased if all response probabilities are equal. In this case, selection problems will only lead to fewer observations, but the composition of the sample is not affected. The idea is to divide the sample in strata in such a way that all elements within a stratum have (approximately) the same response probabilities. Consequently, unbiased estimates can be computed within strata. Next, stratum estimates are combined into a population estimate.

In case of response propensity stratification the final estimates rely less heavily on the correctness of the model that is used to calculate the response propensities. The reason is that not the exact values are used in the computation. They are just used to construct strata.

Suppose the sample is stratified into L strata based on the response propensities. Cochran (1968) suggests that five strata are sufficient, i.e. $L = 5$. The strata are denoted by U_1, U_2, ..., U_L. The sample size in stratum h is denoted by n_h. Assuming simple random sampling, the *response propensity estimator* for the population mean of the target variable Y is now defined by

$$\bar{y}_{RPS} = \frac{1}{n}\sum_{h=1}^{L} n_h \bar{y}_R^{(h)} \tag{16}$$

where $\bar{y}_R^{(h)}$ is the mean of the responding elements in stratum h, for $h = 1, 2, \ldots, L$.

Note that this post-stratification estimator calibrates the response to the sample level instead of the population level (as the post-stratification estimator in section 5.4).

There are a number of ways to construct strata. Preferably, the strata should be constructed in such a way that the response propensities vary as little as possible within strata. Starting point is the distribution of the estimated response propensities. Then decisions have to be made about the number of strata and about the width of the strata (in terms of values of the response propensities).

According to Cochran (1968), five strata are enough for stratification purposes. This is a rule of thumb. However, one should notice that the more strata there are, the less variation there will be within strata.

5.5 Which technique to use?

There are several weighting techniques to choose from. One can apply post-stratification (or one of its extensions, i.e. generalized regression estimation or raking ratio estimation). Another approach is to first estimate response probabilities and then apply one of the techniques based on propensity scores.

Currently no clear statement can be made as to which technique works best to remove or reduce a bias. There is some anecdotal evidence that the effect of weighting adjustment depends more on the auxiliary variables used than on the technique, but more research is necessary.

6. Measurement errors

A measurement error is a discrepancy between the concept that is meant to be measured by a question, and the answer of the respondent to that question. Measurement errors arise either because the question and/or the answering options are not well designed, or because the respondent shortcuts the answering process. Measurement errors are sometimes referred to as response bias, as opposed to nonresponse bias.

Whether unintended or deliberate, during the (cognitive) process of answering questions errors may occur resulting in response bias. The answering process consists of a number of phases (Tourangeau and Rasinski, 1988): Interpretation and comprehension of the question and the response options, information retrieval from memory or other sources, judgement of information to form an answer, reporting the answer in the requested response format.

When it comes to errors that may occur during this process, it is important to distinguish between self-administered and interviewer-assisted survey modes. In interviewer-assisted surveys, the interviewer can assist the respondent with the interpretation and comprehension of the question. In the information retrieval phase, the interviewer can probe the respondent to think harder, or offer options that can help in finding the right information. But the interviewer also controls the speed of the interview, possibly allowing for less time to think than would be needed by the respondent. In the judgement phase, interviewers can motivate respondents to continue when judgement is hard, but their mere presence may lead to a social

desirability bias. This occurs when respondents alter their answer into an answer they think is more social desirable. Social desirable answering behaviour results in overreporting of behaviour that is considered desirable such as voting or health status. It results in underreporting of behaviour that is considered undesirable such as smoking and alcohol comsumption. While reporting, social desirability also plays a role but the interviewer can also assist the respondent in formulating or choosing the answer.

In self-administered questionnaires, the role of the questionnaire design is critical. The persuasive power and aid of the interviewer is absent and the respondent is by himself with only the questionnaire. For this reason, a lot of methodological work is devoted to questionnaire design. See for example the work done by Dillman and Tarnai (1991) and Dillman (2000). Pre-testing Web questionnaires is an important component of this work. Techniques that can be used for this are for instance eye-tracking, qualitative interviews, or experiments in a questionnaire lab.

The absence of the interviewer leads to a lower response rate, but usually also to less social desirable answers (Baker et al., 2010). In addition, the amount of satisficing tends to be lower in self-administered surveys (Baker et al., 2010). Satisficing was first introduced by Krosnick (1991) and implies that respondents do not fully and comprehensively follow the four phases of the answering process, leading to erroneous answers. There are many forms of satisficing, which may result in different types of response behaviour. Examples are: choosing the first option that seems reasonable, or non-differentiation. Ariel et al. (2008) present an overview from the literature. Satisficing is induced by decreased motivation of the respondent, and by a high (perceived) burden of the question.

In self-administered surveys, the ‘don’t know’ answer option is explicitly presented to the respondent. Therefore, this option is more often chosen than in interviewer-assisted surveys where the interviewer has to judge whether someone doesn’t know the answer to a question. Generally, more don’t know-answers are assumed to be of a lower data quality. Whether this is true is unclear. For factual questions, it may be true since the answer to those questions is probably available somewhere and just needs to be looked up. For attitudinal questions, however, it could be not true as persons really may not have thought about certain issues before the survey and hence truly not know what they think about something.

The causes for measurement error described above apply for all surveys, whether cross-sectional, in a panel, unimode or mixed-mode. In addition, there is one phenomenon that causes measurement error specifically in panel studies. This is *panel conditioning*: when answering many surveys influences the respondents’ behaviour or attitudes. Participating in online panels is believed to have this effect because respondents answer more surveys than respondents that are selected for one single survey, and because respondents to online panels typically participate in more than one panel. The latter holds mostly for nonprobability panels. However, the literature does not present clear findings on panel conditioning effects (Baker et al., 2010). Toepoel, Das and van Soest (2008) compared the answering behaviour of

more experienced panel members to less experienced members and found few differences.

The risk of panel conditioning is less in a probability based Web panel. Furthermore, by varying the topic of the survey from wave to wave panel conditioning can be controlled (see e.g. Dennis 2001).

Ideally, differences in measurement between modes should be dealt with in questionnaire design. However, some measurement effects will always remain. It is then the question what is the true answer. Measurement differences can be analysed by asking similar questions in a reference survey. The method of propensity scores can be used to control for differences in selection in the different surveys. The remaining difference can - under the assumption that the method accounts for all differences due to selection - be ascribed to differences in response patterns. The techniques presented in the previous section can be applied in a slightly different way to analyse this (see also Lugtig et al., 2011).

7. Discussion

Web panels may seem a cheap mode to collect data, compared to more traditional data collection modes. However, as is outlined in this paper there are many methodological issues that need to be tackled to make statistics for the general population of an acceptable quality with a web panel. Solving these issues in a satisfying way will be costly.

First, it needs to be decided what type of panel is desirable; a longitudinal panel, a cross-sectional panel or a combination of both. The type of panel sets the conditions for other issues in the implementation. One of these issues is the recruitment of a representative pool of respondents. Dealing with recruitment nonresponse and attrition can be done in such a way that the panel is representative for some pre-defined characteristics available in the sampling frame. In addition, adjustment methods can be used to account for any bias due to selective nonresponse. A lot of effort should be directed at the recruitment of a representative pool of respondents, and involving them in the panel when they have just joined. Literature shows that once a member becomes active, the probability of drop out is less high. It seems that the use of interviewer-assisted modes in the recruitment of a panel work best and that incentives largely increases the response rate. But any decision on the recruitment strategy also relies heavily on a costs-quality assessment. Therefore what the best way is to recruit a panel, and how persons can be motivated to participate and stay active in the panel needs to be evaluated in practice.

Another issue that needs to be tackled is how a panel can best be maintained. What should be the frequency of the measurements? This can only be determined when taking into account other aspects, like the variability of the behaviour that is being measured, the duration of the surveys, the perceived response burden, the rotation scheme and the maximum duration.

Nonresponse during recruitment, in specific waves as well as drop out will lead to bias. Recruitment nonresponse should be minimised by choosing the best recruitment method. Always some recruitment nonresponse will remain. This can be handled with adjustment methods described in section 4. To deal with attrition, substantive information from previous waves can be used to assess and adjust for attrition bias. It is also possible to use information from a reference survey to assess the quality of the data from the web panel and possibly even to adjust for bias.

Measurement effects occur in every survey where a questionnaire is used. Measurement effects can be caused by the assistance of an interviewer, or the lack of assistance of an interviewer. Especially questions that are very personal, sensitive, that measure social undesirable behaviour or for which the respondent may need some time to think are very suitable for a web questionnaire. However, due to the absence of an interviewer, questions should not be too difficult as this may lead to drop out. To analyse measurement effects, a reference survey can be used that asks the exact same questions. The propensity score method can be applied to separate selection effects and measurement effects. However, it still remains the question which survey measured the real value.

Even if these issues are solved in an acceptable way, then a Web panel will still not be able to replace the traditional survey modes CAPI and CATI. It is, however, possible to identify topics / questions that can be asked in a web panel without a loss of quality. Maybe even of better quality. In addition, the speed and longitudinal aspect of a web panel enable the collection of more timely and hence topical information and monitor (changes in) the behaviour of respondents.

The Internet is becoming more and more common in present society, creating dozens of possibilities to facilitate all aspects of everyday life. Take for example a not even very recent development like the introduction of smart phones and applications (apps) for smartphones. Internet penetration rates are as high as 98% in The Netherlands, and increasing. The part of the population that is not familiar with the use of the Internet is ageing whereas new generations grow up surrounded by tablets, smartphones and Webshops.

At the same time, the percentage of persons that can be reached by a registered, fixed-line telephone is decreasing. Face-to-face visits are becoming more exceptional due to the increase in contact over the Web and the rising costs per response. In addition, the Internet has made it easier for other parties to collect data. As we live in an information society, this puts pressure on the position of the National Statistical Office as the authority on data collection for statistical information on the society.

These arguments are all in favour of investing in research and practices for making web panels fit for use for official statistics. In this paper, a number of methodological issues have been identified that need to be resolved before a web panel can be implemented for the production of official statistics. For some of these issues, suggestions on how these should be dealt with are provided in the paper. For other issues, experimental research is needed to guide methodological decision

making. In addition, not all the issues are equally important. This depends on whether the costs, quality or a combination or both is of main interest.

A summary of these issues is provided below, with some suggestions. Experimental research will need to provide the insights in each of these issues to identify the most desirable web panel strategy for official statistics.

1. What are the effects of different recruitment strategies on recruitment response and attrition in subsequent waves?
    - How can the reasons for participation in a web panel be used to increase response (advance letters, recruitment calls etc.)?
    - What mode or combination of modes is best for recruitment? *One that covers the entire population. Take into account the costs of recruitment.*
    - Use of refusal conversion for the recruitment interview?
    - How to deal with undercoverage: providing Internet for persons without access?
    - What questions should be asked in the recruitment interview? *Attitudinal questions (MEPS), webograhics, psychographics. Questions from the core questionnaire of Dutch household surveys for comparison and adjustment (propensity method).*
    - Can an existing survey be used to recruit panelists?
    - Should incentives be used, and how? For recruitment, for specific groups, depending on survey duration? *LISS-panel: Pre-paid incentive of 10 euros increased the recruitment response with almost 25%.*
2. How should the web panel be maintained so that panel conditioning effects and attrition bias are kept to a minimum:
    - What topics are fit for use in a web panel? *Vary the topic to reduce the risk of panel conditioning.*
    - And what type of questions? *Especially questions that are very personal, sensitive, that measure social undesirable behaviour or for which the respondent may need some time to think are very suitable for a Web questionnaire. However, due to the absence of an interviewer, questions should not be too difficult as this may lead to drop out.*
    - What should be the frequency and the duration of the panel surveys? *Frequency: once or twice a month, duration: 10 minutes, 15 max.*
    - What is the optimal rotation scheme and maximum duration of panel membership? *Maximum duration: one year.*
3. How to keep (especially new) panelists involved with the panel?

4. What to do with sleepers, use of refreshment samples? *Use refreshment samples to selectively replace attriters.*
5. What is the best adjustment strategy for recruitment nonresponse and panel attrition? *Deal with recruitment nonresponse and attrition separately.*
6. How to deal with measurement errors? *Handle these in the questionnaire design: pre-testing. Use a reference survey to analyse differences in measurement.*

8. References

American Association for Public Opinion Research, AAPOR (2006), *Final Dispositions of Case Codes and Outcomes Rates for Surveys, 4th Edition*. AAPOR, Lenexa, KS, USA

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

Bethlehem, J.G. (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. In: Groves, R.M., Dillman, D.A., Eltinge, J.L. & Little, R.J.A. (eds), *Survey Nonresponse* (eds.), John Wiley & Sons, New York

Bethlehem, J.G., Cobben, F. & Schouten, B. (2011), *Handbook of Nonresponse in Household Surveys*. John Wiley & Sons, Hoboken, NJ

Börsch-Supan, A., Elsner, D., Faßbender, H., Kiefer, R., McFadden, D. & Winter, J. (2004), *Correcting the Participation Bias in an Online Survey*. Report, University of Munich, Munich, Germany

Callegaro, M. & DiSogra, C.(2008). Computing Response Metrics for Online Panels, *Public Opinion Quarterly* 72, pp. 1008–1032

Cialdini, R. (1988), Influence: Science and Practice, Glenview IL: Scott, Foresman and Co

Cochran, W.G. (1968), The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. *Biometrics* 24, pp. 205 – 213

Cochran, W.G. (1977). *Sampling Techniques*, Third Edition. New York: John Wiley & Sons, New York

Couper, M. (2007). Issues of Representation in Health Research (with a Focus on Web Surveys). *American Journal of Preventive Medicine* 32, pp. 83–89

Das, M., V. Toepoel and A. van Soest (2011), Nonparametric Tests of Panel Conditioning and Attrition Bias, *Sociological Methods and Research*, vol 40 (1), pp. 32-56

Dillman, D. (2000), Mail and Internet Surveys: The Total Design Method. New York: Wiley and Sons, Inc.

Dillman, D., and J. Tarnai (1991), Mode Effects of Cognitive Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys. In: Measurement Errors in Surveys, edited by P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman. J. Wiley & Sons, New Jersey.

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

Groves, R.M. and Couper, M.P. (1998), *Nonresponse in Household Interview Surveys*, Wiley, New York.

Groves, R.M. Singer, E. and Corning, A. (2000), Leverage-saliency theory of survey participation: Description and illustration. Public Opinion Quarterly, 64, pp. 299 - 308

Hansen, K.M. and Pedersen, R.T. (2012), Efficiency of Different Recruitment Strategies for Web panels, International Journal of Public Opinion Research, Vol. 24, No. 2, pp. 238 – 249

Horvitz, D.G. & Thompson, D.J. (1952), A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47, pp. 663-685

Huggins, V. & Eyerman J. (2001), *Probability Based Internet Surveys: A Synopsis of Early Methods and Survey Research Results*. Paper presented at Federal Committee on Statistical Methodology Research Conference, November 14–16, Arlington, VA, USA

Laurie, H., and Lynn, P. (2008), The Use of Respondent Incentives on Longitudinal Surveys, Paper No. 2008-42, Institute for Social and Economic Research, UK

Lugtig, P., Lensveldt-Mulders, G., Frerichs, R. and Greven, F. (2011), Estimating nonresponse bias and mode effects in a mixed-mode survey, International Journal of Market Research, Vol 53, Issue 5, pp. 669 – 686

Office of Management and Budget, OMB (2006). *Questions and Answers When Designing Surveys for Information Collections*. Office of Management and Budget, Washington, DC, USA

Rao, K., Kaminska, O. and McCutcheon, A.L. (2010), Recruiting probability samples for a multi-mode research panel with internet and mail components, Public Opinion Quarterly, 74, pp. 68-84

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

Scherpenzeel, A. and 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

Scherpenzeel, A. (2009), Innovations and New Technologies in Panel Research, paper presented at the 5th International conference on panel data users, Lausanne, Switzerland

Schlengen, W.E., Caddell J. M., Ebert L., Jordan K.B., Rourke K.M., Wilson D., Thalji T., Dennis M.J., Fairbank J. A. and Kulka, R. (2002), Psychological Reactions to Terrorist Attacks. Findings from the National Study of American's Reactions to September 11. *Journal of the American Medical Association* 288, pp. 581–588

Schonlau M., Van Soest A, Kapteyn A, and Couper M. (2009). Selection Bias in Web Surveys and the Use of Propensity Scores. *Sociological Methods and Research* 37, pp. 291-318

Schonlau, M., Zapert, K., Payne Simon, L., Haynes Sanstad, K., Marcus, S., Adams, J. Kan, H., Turber, R. & Berry, S. (2004), A Comparison Between Responses from Propensity-weighted Web Survey and an Identical RDD Survey. *Social Science Computer Review* 22, pp. 128-138

Sikkel, D. & Hoogendoorn A. (2008). Panel Surveys. In: De Leeuw, E., Hox, J., & Dillman D.A. (Eds.), *International Handbook of Survey Methodology*, Lawrence Erlbaum. New York, USA, pp. 479–499

Sikkel, D. and Hoogendoorn, A. (2008), Panel Surveys, Chapter 25 in: International Handbook of Survey Methodology, edited by E. De Leeuw, J. Hox and D. Dillman

Tourangeau, R. (2003). Web-Based Data Collection. In: Cork D.L., Cohen, M.L., Groves, R. & Kalsbeek, W. (Eds), *Survey Automation, Report and Workshop Proceedings*, National Academies Press, Washington, DC, USA. pp. 183–198

Tourangeau, R. and Rasinski, K.A. (1988), Cognitive processes underlying context effects in attitude measurement, Psychological Bulletin, 103, 299-314

Van Ossenbruggen, R., Vonk, T, and Willems, P. (2006), Results Dutch Online Panels Comparison Study (NOPVO), source: www.nopvo.nl